



**Sara Maria Rodrigues Mineiro dos Santos Luís**

Licenciatura em Engenharia Informática

## **Aplicação Móvel para Análise de Biossensores Diversificados**

Dissertação para obtenção do Grau de Mestre em  
**Engenharia Informática**

Orientador: Nuno Manuel Robalo Correia,  
Professor Catedrático,  
Universidade Nova de Lisboa

Co-orientadora: Elvira Maria Correia Fortunato,  
Professora Catedrática,  
Universidade Nova de Lisboa

Júri

Presidente: Doutor José Augusto Legatheaux Martins  
Vogais: Doutor Manuel João Caneira Monteiro da Fonseca  
Doutor Nuno Manuel Robalo Correia



FACULDADE DE  
CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE NOVA DE LISBOA

**Dezembro, 2019**



## **Aplicação Móvel para Análise de Biossensores Diversificados**

Copyright © Sara Maria Rodrigues Mineiro dos Santos Luís, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.





## AGRADECIMENTOS

Quero agradecer em primeiro lugar aos meus orientadores, o Professor Doutor Nuno Correia e a Professora Doutora Elvira Fortunato, pelo apoio, sugestões e disponibilidade durante a elaboração deste trabalho.

Gostaria de agradecer também ao Tomás Pinheiro do Centro de Investigação em Materiais (CENIMAT) pelo entusiasmo sempre presente em relação ao projeto e pela disponibilidade que sempre demonstrou para explicar quaisquer questões relacionadas com os biossensores ou para fornecer novas amostras quando necessárias.

Agradeço ainda ao André Rodrigues por todo o apoio e pelo tempo infindável que perdeu a responder às minhas questões sobre o projeto. A todos os colegas do departamento de informática que fizeram comigo este percurso ao longo de cinco anos e em especial à Sílvia, à Ana, ao António, aos Andrés, ao João, à Khrystyna, ao Tiago e ao Henrique, obrigada por tudo.

Por último, agradeço a toda a minha família pela paciência e pelo apoio incondicional em tudo o que faço.



## RESUMO

---

O rápido aumento da prevalência de doenças não transmissíveis como o cancro, a diabetes e as doenças cardiovasculares, em países com poucos meios económicos constitui um importante problema de saúde pública mundial. Nestas situações, para que haja uma redução do número de casos e para aumentar a qualidade de vida dos doentes, o diagnóstico precoce é essencial. No entanto, nestes países o acesso a meios de diagnóstico é muito limitado, seja pelo elevado custo dos mesmos ou pela falta de acessibilidade ao tratamento por parte das populações. Existe então a necessidade de desenvolver meios de diagnóstico alternativos que sejam simples, portáteis, acessíveis e de baixo custo.

Uma potencial solução consiste no uso do papel para a criação de dispositivos de diagnóstico pois este é um material amplamente disponível e de baixo custo. Neste âmbito foi criado um biossensor colorimétrico em papel que permite a leitura simultânea da concentração de glucose, ácido úrico, pH e colesterol, através da observação de uma mudança de cor. Substâncias essas que são consideradas relevantes para o diagnóstico e prevenção de problemas como a diabetes e as doenças do fígado, dos rins e cardiovasculares. Para complementar o sensor foi criada, em parceria com o CENIMAT, uma aplicação móvel que realiza a análise da cor através da captura de imagens do mesmo, apresentando um resultado que permite fazer um diagnóstico. Esta conjugação de tecnologias permite realizar uma análise rápida e precisa num dispositivo portátil e acessível a todos.

O objetivo deste trabalho é a adaptação dessa aplicação móvel para a deteção e análise colorimétrica do novo sensor, aplicando métodos de processamento de imagem e aprendizagem automática para obter resultados relativos à glucose e ao ácido úrico. Assim, foram criados modelos de aprendizagem automática que classificam estas duas substâncias de acordo com níveis de referência. Para complementar a análise foi desenvolvido um método de calibração da cor do sensor que reduz o efeito de influências exteriores nos resultados da análise.

**Palavras-chave:** Dispositivos Móveis, Processamento de Imagem, Biossensor Colorimétrico, Análise Colorimétrica, Classificação, Calibração de cor, Papel, Glucose, Ácido Úrico

---



## ABSTRACT

---

The rapid increase in the prevalence of non-communicable diseases such as cancer, diabetes and cardiovascular disease, in countries with few economic resources is a major global public health concern. In these situations, early diagnosis is essential to reduce the number of cases and increase the quality of life of the patients. However, the access to diagnostic materials in these countries is very limited, either because of their high cost or because of the lack of accessibility to treatment for the population. There is therefore a need to develop alternative diagnostic means that are simple, portable, widely accessible and inexpensive.

A potential solution to this problem is the use of paper as a substrate for the creation of diagnostic devices because it is a widely available and low-cost material. In this context, a paper-based colorimetric biosensor was created to simultaneously read the concentration of glucose, uric acid, pH and cholesterol in a sample by observing a colour change. These substances are considered relevant markers for the diagnosis and prevention of problems such as diabetes and liver, kidney and cardiovascular diseases.

To complement the sensor, an application for mobile devices was created in partnership with CENIMAT. This application captures images of the sensor, performs the colour analysis and presents a result that provides a diagnosis to the user. This combination of technologies enables rapid and accurate analysis on a portable device that is accessible to everyone.

The main objective of this work is the adaptation of this mobile application for the detection and colorimetric analysis of the new sensor, applying image processing methods and machine learning to obtain diagnosis results related to glucose and uric acid. Thus, automatic learning models were created that classify these two substances according to reference levels. To complement the analysis, a method was also developed to calibrate the colour and reduce the effect of external influences on the analysis results.

**Keywords:** Mobile Devices, Image Processing, Colorimetric Biosensor, Colorimetric Analysis, Classification, Colour Calibration, Paper, Glucose, Uric Acid

---



# ÍNDICE

<b>Siglas</b>	<b>xv</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	1
1.2 Descrição e Contexto . . . . .	2
1.2.1 Contexto do Projeto . . . . .	2
1.2.2 Descrição do Problema . . . . .	3
1.3 Principais Objetivos e Contribuições . . . . .	4
1.4 Organização do Documento . . . . .	5
<b>2 Enquadramento</b>	<b>7</b>
2.1 Processamento de Imagem . . . . .	7
2.1.1 Cor . . . . .	7
2.1.2 Descritores de Características da Imagem . . . . .	10
2.1.3 Classificação . . . . .	13
<b>3 Trabalho Relacionado</b>	<b>19</b>
3.1 Análise Colorimétrica . . . . .	19
3.1.1 Observação de Curvas de Calibração . . . . .	19
3.1.2 Abordagens com recurso a Aprendizagem Automática . . . . .	21
3.1.3 Versão Anterior da Aplicação . . . . .	24
3.1.4 Análise Crítica . . . . .	25
3.2 Calibração da Cor . . . . .	26
3.2.1 Acessórios para o <i>Smartphone</i> . . . . .	26
3.2.2 Métodos de Compensação . . . . .	27
3.2.3 Versão Anterior da Aplicação . . . . .	34
3.2.4 Análise Crítica . . . . .	34
<b>4 Métodos</b>	<b>37</b>
4.1 Análise Colorimétrica . . . . .	37
4.1.1 Obtenção e Processamento de Dados . . . . .	38
4.1.2 Classificação . . . . .	45
4.2 Calibração da Cor . . . . .	48

4.2.1	Modificação do Algoritmo Usado na Versão Anterior . . . . .	49
4.2.2	Captura e Correção da Imagem . . . . .	49
4.2.3	Comparação da Cor . . . . .	51
<b>5</b>	<b>Resultados</b>	<b>53</b>
5.1	Análise Colorimétrica . . . . .	53
5.1.1	Classificação da Glucose . . . . .	54
5.1.2	Classificação do Ácido Úrico . . . . .	59
5.2	Calibração . . . . .	61
5.2.1	Seleção do Algoritmo a Implementar . . . . .	62
5.2.2	Alterações no Biossensor . . . . .	64
<b>6</b>	<b>Implementação</b>	<b>67</b>
6.1	Deteção do Sensor . . . . .	67
6.2	Calibração da Cor . . . . .	71
6.3	Análise Colorimétrica . . . . .	73
6.3.1	Implementação . . . . .	73
6.3.2	Procedimento da Análise . . . . .	74
6.3.3	Testes . . . . .	74
<b>7</b>	<b>Conclusões e Trabalho Futuro</b>	<b>79</b>
7.1	Trabalho Futuro . . . . .	80
	<b>Bibliografia</b>	<b>83</b>
<b>A</b>	<b>Resultados da Comparação da Cor</b>	<b>89</b>
A.1	Comparação da Correção das Imagens sob os Três Tipos de Iluminação Analisados . . . . .	89
A.2	Comparação da Correção das Imagens entre os Modelos de <i>Smartphone</i> Analisados . . . . .	90
A.3	Comparação entre os Resultados da Correção e as Referências . . . . .	90
<b>B</b>	<b>Resultados da Correção da Cor</b>	<b>93</b>
B.1	Resultados da Correção para Comparação dos Algoritmos de Calibração Estudados . . . . .	93
B.2	Outros Sensores Usados na Comparação . . . . .	126
B.3	Resultados da Correção Através da Aplicação Móvel . . . . .	126
<b>C</b>	<b>Resultados da Classificação do Ácido Úrico</b>	<b>129</b>
C.1	Modelos RGB . . . . .	130
C.2	Modelos HSV . . . . .	131
C.3	Modelos XYZ . . . . .	132
C.4	Modelos HLS . . . . .	133



C.5	Modelos LAB . . . . .	134
C.6	Modelos YUV . . . . .	135
C.7	Rácio R/B . . . . .	136
C.8	Componente H do Modelo HSV . . . . .	136
C.9	Rácio R/B e Componente H do Modelo HSV . . . . .	137
C.10	Componentes A e B do Modelo LAB . . . . .	138
C.11	Componentes H e S do Modelo HSV . . . . .	139
C.12	Componentes H e S do Modelo HLS . . . . .	140
C.13	Componentes H e S dos Modelos HSV e HLS . . . . .	141
C.14	Modelo RGB, Componente V (HSV), Componente Z (XYZ) , Componente A (LAB) e Componente Y (YUV). . . . .	142
<b>D</b>	<b>Resultados da Classificação da Glucose</b>	<b>143</b>
D.1	Modelos Com a Média de RGB . . . . .	144
D.2	Modelos Com a Média de HSV . . . . .	146
D.3	Modelos Com a Média de XYZ . . . . .	148
D.4	Modelos Com a Média de HLS . . . . .	150
D.5	Modelos Com a Média de LAB . . . . .	152
D.6	Modelos Com a Média de YUV . . . . .	154
D.7	Modelos Com a Média, Desvio padrão e Assimetria RGB . . . . .	156
D.8	Modelos Com a Média, Desvio padrão e Assimetria HSV . . . . .	158
D.9	Modelos Com a Média, Desvio padrão e Assimetria XYZ . . . . .	160
D.10	Modelos Com a Média, Desvio padrão e Assimetria HLS . . . . .	162
D.11	Modelos Com a Média, Desvio padrão e Assimetria LAB . . . . .	164
D.12	Modelos Com a Média, Desvio padrão e Assimetria YUV . . . . .	166
D.13	Rácio R/G . . . . .	168
D.14	Média dos Componentes V e S do Modelo HSV . . . . .	169
D.15	Média dos Componentes L e S do Modelo HLS . . . . .	171
<b>E</b>	<b>Histogramas</b>	<b>175</b>
E.1	Glucose . . . . .	175
E.2	Ácido Úrico . . . . .	178
<b>F</b>	<b>Regiões de Decisão dos Modelos Desenvolvidos</b>	<b>183</b>
F.1	Concentrações de Glucose . . . . .	184
F.2	Quatro Níveis de Glucose . . . . .	199
F.3	Cinco Níveis de Glucose . . . . .	229
F.4	Concentrações de Ácido Úrico . . . . .	259
F.5	Níveis de Ácido Úrico . . . . .	273
<b>G</b>	<b>Resultados da Análise de Componentes Principais (PCA)</b>	<b>287</b>
G.1	Concentrações de Glucose . . . . .	287

## ÍNDICE

---

G.2	Quatro Níveis de Glucose . . . . .	293
G.3	Cinco Níveis de Glucose . . . . .	303
G.4	Concentrações de Ácido Úrico . . . . .	313
G.5	Níveis de Ácido Úrico . . . . .	317
<b>H</b>	<b>Interface da Aplicação</b>	<b>321</b>

## SIGLAS

CCV	<i>Color Coherent Vector.</i>
CENIMAT	Centro de Investigação em Materiais.
CIE	Comissão Internacional de Iluminação.
FCT-UNL	Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa.
HCA	<i>Hierarchical Cluster Analysis.</i>
ISO	Organização Internacional de Normalização.
KNN	<i>K-Nearest Neighbours.</i>
LS-SVM	<i>Least Squares Support Vector Machines.</i>
NFC	<i>Near Field Communication.</i>
NOVA LINCS	<i>NOVA Laboratory for Computer Science and Informatics.</i>
PCA	Análise de Componentes Principais.
POCT	Point of Care Testing.
SIFT	<i>Scale Invariant Feature Transform.</i>
SVM	<i>Support Vector Machines.</i>



## INTRODUÇÃO

Neste capítulo é feita uma breve descrição da motivação e do contexto em que se enquadra esta dissertação. São também apresentados os objetivos e principais contribuições previstas para o desenvolvimento deste projeto.

### 1.1 Motivação

Atualmente, as doenças não transmissíveis são a maior causa de morte em todo o mundo. Em 2016, problemas como o cancro, a diabetes e as doenças cardiovasculares foram responsáveis por 71% das 57 milhões de mortes ocorridas no planeta, sendo que 78% desses casos ocorreram em países de baixo e de médio rendimento [30]. O crescente nível de vida em alguns destes países, como o Brasil ou a China, levou a um aumento considerável do número de indivíduos afetados por estas doenças [29].

Para haver uma redução do número de casos e para o aumento da qualidade de vida dos doentes, o diagnóstico precoce é essencial [28]. Por este motivo, o acesso a meios de diagnóstico é vital e deve estar disponível para todas as populações afetadas.

Normalmente, as análises necessárias são efetuadas em laboratório com recurso a equipamento e pessoal especializado. No entanto, nos países em vias de desenvolvimento há grandes disparidades no acesso a cuidados de saúde, sendo que os indivíduos sujeitos a desvantagens socioeconómicas, particularmente os que residem em meios remotos, têm grande dificuldade em aceder a quaisquer meios de diagnóstico [29].

Consequentemente, há uma tendência para o desenvolvimento de dispositivos de diagnóstico alternativos, nomeadamente para Point of Care Testing (POCT). Estes testes de diagnóstico, realizados junto do paciente, são de baixo custo, pouco invasivos, precisos e permitem obter um resultado rápido [49].

Uma potencial solução para este problema consiste no uso do papel como substrato

para a criação de biossensores de diagnóstico, pois este é um material amplamente disponível e de baixo custo, tendo já sido demonstrada a sua capacidade para suportar a deteção de várias substâncias a partir da observação de reações colorimétricas [12, 22, 53].

Os biossensores microfluídicos baseados em papel tem um baixo custo, não requerem equipamento específico e são de fácil utilização. Ao serem *multiplex* e não enzimáticos trazem outro conjunto de vantagens. Para além de permitirem a análise de várias substâncias em simultâneo, são mais resistentes a condições ambientais e de armazenamento. Estes fatores tornam estes dispositivos particularmente importantes para aplicação em regiões isoladas com pouca acessibilidade.

Para obter um sistema de diagnóstico completo, para além do biossensor são necessários dispositivos acessíveis e de baixo custo que consigam fazer a sua leitura e apresentar o resultado. Os *Smartphones* estão em constante evolução e têm cada vez mais elevadas capacidades de processamento e de memória e ainda câmaras de alta resolução. Cerca de 94% da população mundial tem acesso a um *smartphone*, sendo que 70% desses indivíduos residem em países em desenvolvimento [48]. Estes fatores tornam estes dispositivos extremamente acessíveis e com capacidade para realizar a análise destes biossensores.

Assim, a conjugação de sensores de baixo custo com um meio de análise amplamente disponível e portátil como o *smartphone*, constitui um sistema de diagnóstico completo. É potencialmente útil para melhorar o alcance dos sistemas de saúde, fornecendo serviços de proximidade a populações mais remotas e com poucos recursos [48].

## 1.2 Descrição e Contexto

### 1.2.1 Contexto do Projeto

A presente dissertação surge enquadrada num projeto de colaboração entre o Departamento de Informática da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (FCT-UNL), o *NOVA Laboratory for Computer Science and Informatics* (NOVA LINCS) e o CENIMAT, mais precisamente na continuação do trabalho desenvolvido noutra dissertação de Mestrado realizada por Rodrigues [37, 38]. O propósito final desse trabalho foi o desenvolvimento de uma aplicação móvel capaz de realizar a análise colorimétrica de um biossensor não enzimático em papel, através de algoritmos de aprendizagem automática. Esse biossensor foi proposto por Ferrão [8] e permite a determinação do valor de diabetes a partir da concentração de glucose no sangue, tirando partido da tecnologia *Lab-on-Paper* que se baseia em reações químicas sobre papel.

A aplicação desenvolvida por Rodrigues [37] realiza a análise colorimétrica da concentração de glucose a partir de uma imagem do biossensor. Após a captura da imagem através da câmara do *smartphone*, procede-se à deteção das zonas do sensor relevantes para a análise, através de vários métodos de processamento de imagem. Com essas zonas identificadas realiza-se a calibração da cor da imagem através da aplicação de uma matriz

de correção. Após a correção da imagem, realiza-se a análise da cor. Para fazer a análise da cor foram desenvolvidos dois classificadores, capazes de estabelecer uma correspondência entre as cores das amostras e as concentrações que estas representam, de modo a apresentar um resultado que possibilite o diagnóstico.

Foram apresentados resultados de precisão e desempenho elevados numa análise qualitativa, o que está de acordo com as características do sensor que permitem distinguir entre concentrações reduzidas e elevadas com qualidade. Em relação à análise quantitativa, não foi possível fazer uma distinção muito eficaz das concentrações de glucose. Esta análise é particularmente importante para utilizadores que precisam de controlar a diabetes com maior detalhe. Considerou-se ainda que as funcionalidades desenvolvidas têm potencial para ser aplicadas em contexto clínico.

Para melhorar os resultados foram propostos o aumento do conjunto de dados utilizado para treinar os classificadores, bem como melhorias na calibração de cor do sensor para que seja possível obter uma cor menos influenciada por fatores externos. Foram ainda sugeridas melhorias na gama de cores usada pelo biossensor, para que esta apresentasse cores mais distinguíveis entre as várias concentrações.

### 1.2.2 Descrição do Problema

A presente dissertação surge da necessidade de continuar a desenvolver a aplicação móvel para a análise de um novo sensor, proposto por Pinheiro [33]. Neste trabalho foi desenvolvido um biossensor *multiplex* não enzimático feito em papel com a capacidade de medir simultaneamente concentrações de glucose, ácido úrico, colesterol e valores de pH.

Este dispositivo tira partido da tecnologia *Lab-On-Paper*, que recorre ao papel como substrato para as reações químicas, utilizando impressão a cera para criar canais de microfluídica. A análise é baseada na observação das propriedades colorimétricas da reação de uma determinada substância com nano partículas de ouro. Através da realização de testes com soluções de diferentes concentrações das substâncias a analisar, seguidos de uma análise digital, foi possível desenvolver curvas de calibração que permitem extrapolar as suas concentrações.

As substâncias presentes neste biossensor em particular são biomarcadores muito importantes, cuja medição permite detetar e prevenir doenças como a diabetes e as doenças do fígado, rins ou cardiovasculares. O dispositivo é então dirigido a indivíduos sujeitos a estes problemas, principalmente em países em desenvolvimento, por se tratar de um sistema de baixo custo e fácil utilização. O uso destes materiais e reações não enzimáticas dá ao sensor uma maior resistência às condições ambientais e de armazenamento, o que é uma vantagem competitiva face a outros dispositivos semelhantes. A junção deste biossensor com um método de leitura acessível da reação colorimétrica, como a análise de imagem através de um *smartphone*, dá origem a um sistema de diagnóstico completo, portátil e de baixo custo.

Apesar de presentes no biossensor, o colesterol e o pH não são alvo de estudo nesta dissertação. O CENIMAT considerou que a reação colorimétrica existente para o colesterol ainda não era suficiente para permitir obter amostras favoráveis para o treino dos classificadores. Por outro lado, o pH não seria relevante como informação para o utilizador, apenas como prova de conceito para a análise colorimétrica através da aplicação.

Assim, com base nos trabalhos propostos anteriormente, pretende-se continuar a desenvolver a aplicação móvel para a análise colorimétrica do novo biossensor. O sistema deve detetar todas as áreas de cor existentes no novo sensor e realizar a análise dos marcadores da glucose e ácido úrico, aplicando um processo semelhante ao utilizado no desenvolvimento da aplicação anterior.

Procura-se também melhorar o processo da análise colorimétrica através da introdução de outras formas de classificação e principalmente da utilização de mais exemplos para treinar os modelos já desenvolvidos. O outro foco do projeto é o desenvolvimento de métodos de calibração da cor para resolver os problemas decorrentes da influência de fatores externos na cor como a luz ambiente e o uso de diferentes modelos de *smartphone*. Estes problemas afetam de forma significativa o resultado da análise, podendo induzir a resultados incorretos.

### 1.3 Principais Objetivos e Contribuições

Espera-se como resultado da dissertação o desenvolvimento de uma aplicação para dispositivos móveis que, através da análise da imagem do biossensor, consiga medir com maior precisão o valor dos diabetes do utilizador a partir da glucose bem como fornecer informação sobre o ácido úrico. Esta será baseada na adaptação do sistema desenvolvido anteriormente [37] a um novo biossensor, sendo descritas de seguida as características esperadas:

- **Adaptação ao novo sensor** – Pretende-se adaptar o algoritmo de deteção já existente à configuração do novo biossensor e a quaisquer modificações no seu *design* necessárias para os processos de calibração da cor desenvolvidos.
- **Calibração da cor** – A calibração da imagem obtida pelo *smartphone* para minimizar a influência da iluminação nos resultados da análise é um passo essencial para melhorar os resultados obtidos na aplicação anterior. Pretende-se estudar formas de resolver este problema, de modo a poder propor uma solução com melhores resultados que possa ser incorporada neste sistema e em futuras utilizações do biossensor.
- **Análise Colorimétrica** – Dada a existência de uma substância adicional face à versão anterior da aplicação, o ácido úrico, é necessário reproduzir o processo de análise e teste realizado anteriormente para a glucose de modo a encontrar modelos que permitam uma classificação correta deste novo marcador. Procura-se ainda



reproduzir o processo para melhorar os resultados obtidos com a glucose através da adição de mais amostras ao conjunto de dados.

- **Interface Gráfica** – O sistema de diagnóstico desenvolvido deve ser fácil de usar para o utilizador, sendo assim importante que a interface seja simples e de adaptação rápida. Como a aplicação apresentou bons resultados nos testes de usabilidade realizados no trabalho anterior, procura-se fazer as alterações necessárias para a análise do novo biossensor com o mínimo de modificações possível ao *design* já existente.

## 1.4 Organização do Documento

O presente documento segue a estrutura apresentada:

- **Capítulo 1 - Introdução:** Esta secção pretende esclarecer o contexto e a motivação para a dissertação e ainda definir as principais contribuições da mesma.
- **Capítulo 2 - Enquadramento:** Neste capítulo é feita uma breve descrição de conceitos teóricos relevantes para o desenvolvimento da dissertação.
- **Capítulo 3 - Trabalho Relacionado:** Neste capítulo é feito um estudo de projetos já realizados relacionados com os vários temas abordados na dissertação, sendo também descritos os processos utilizados na versão anterior da aplicação.
- **Capítulo 4 - Métodos:** Este capítulo contém uma descrição dos processos realizados para o estudo da calibração da cor e da análise colorimétrica.
- **Capítulo 5 - Resultados:** Nesta secção são detalhados os resultados obtidos no estudo da calibração e da análise colorimétrica.
- **Capítulo 6 - Implementação:** Nesta secção é descrita a implementação das soluções encontradas a partir dos resultados obtidos, na aplicação móvel.
- **Capítulo 7 - Conclusões e Trabalho Futuro:** No último capítulo são apresentadas as conclusões retiradas de todo o trabalho realizado na dissertação. É também feita uma proposta de trabalho futuro para melhorar o sistema desenvolvido.



## ENQUADRAMENTO

Neste capítulo é feita uma breve apresentação de conceitos relevantes para o desenvolvimento desta dissertação, nomeadamente noções relacionadas com o processamento de imagem e com a aprendizagem automática.

### 2.1 Processamento de Imagem

O processamento de imagem diz respeito aos processos aplicados a uma imagem com o objetivo de a transformar numa representação de dados de formato utilizável para que possa ser usada depois por vários tipos de aplicações. Nesta secção são descritos vários conceitos relacionados com a imagem que nos permitem entender como extrair as informações necessárias para a sua classificação, bem como métodos para realizar essa classificação.

#### 2.1.1 Cor

A cor é a sensação resultante da incidência de radiação eletromagnética da região do espectro visível na retina. A chegada dessa luz aos olhos após ser refletida por um determinado objeto, provoca a percepção de uma determinada cor para esse objeto. Esta percepção está relacionada com o comprimento de onda da luz que este reflete.

Existem três tipos de células fotorrecetoras na nossa retina chamadas cones. Cada uma destas é sensível a um intervalo de comprimento de onda diferente, sendo esses três intervalos próximos das luzes com as cores vermelha, verde e azul [13]. Essas cores têm propriedades aditivas que permitem a sua mistura para a representação de outras cores diferentes[47]. Todas estas características fazem com que seja natural usar três valores numéricos para representar a cor num sistema digital.

Deste modo, foram desenvolvidos vários modelos para representar a cor computacionalmente, sendo de seguida feita uma breve descrição dos mais utilizados e dos referidos nos capítulos seguintes.

### 2.1.1.1 Modelos de Cor

Em 1931 a Comissão Internacional de Iluminação (CIE) criou um modelo que especifica como a informação de uma determinada luz pode ser transformada em três componentes numéricos que representam a cor, através de experiências realizadas com as três cores primárias com vários observadores [35]. Estas experiências revelaram que para certas partes do espectro era necessário adicionar uma quantidade negativa de uma certa componente para obter uma correspondência de cor. Devido ao problema provocado pelos valores negativos, a CIE desenvolveu um novo espaço de cor chamado XYZ, que representa todas as cores com termos positivos, mapeando o eixo dos  $Y$  para a corresponder à luminância [47].

Foram também definidos valores independentes do brilho como as coordenadas cromáticas. Estas são dadas pelas equações 2.1 e a sua soma devolve sempre um, sendo assim possível obter  $z$  através das coordenadas  $x$  e  $y$ , por exemplo. Estas cores podem ser representadas através do diagrama cromático presente na figura 2.1[13].

$$x = \frac{X}{(X + Y + Z)} \quad y = \frac{Y}{(X + Y + Z)} \quad z = \frac{Z}{(X + Y + Z)} \quad (2.1)$$

Apesar do espaço XYZ permitir a separação entre a luminosidade e a cor, não consegue representar diretamente as diferenças da perceção de diferentes cores. Ao calcular a distância entre duas cores, este valor não irá corresponder à distância entre as cores observadas, isto é, pares de cores com a mesma distância entre si podem ter distâncias observadas muito maiores [13]. Foram então criados espaços como o CIE Lab e o CIE

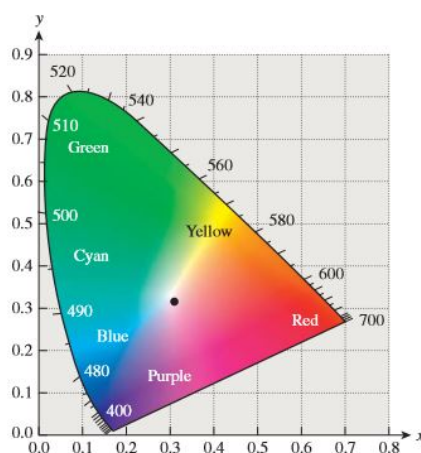


Figura 2.1: Diagrama cromático CIE. A fronteira do diagrama representa a correspondência entre as cores e luzes de um certo comprimento de onda. O ponto no centro representa uma luz branca *standard* chamada "iluminante C"[13].

Luv onde esta relação entre a distância calculada e observada já existe. No CIE Lab,  $L$  representa a luminosidade e o par  $(a, b)$  representa a cor [47].

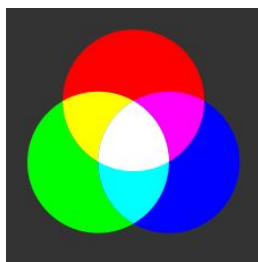
Existem ainda modelos baseados nas propriedades das cores primárias aditivas (vermelho, azul e verde - figura 2.2a) e nas subtrativas (ciano, magenta e amarelo - figura 2.2b). Estes modelos não conseguem definir todas as cores existentes [47]. O modelo aditivo RGB é utilizado na maioria dos monitores. Cada *pixel* é descrito por três números  $(r, g, b)$  que correspondem ao grau com o qual cada uma das três iluminações contribui para a sua cor [13]. O mapeamento do RGB para o CIE XYZ é obtido através da fórmula 2.2 [35].

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.2)$$

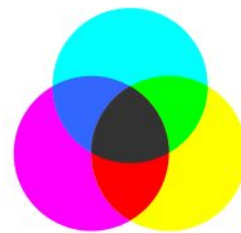
O modelo CMY recorre às cores subtrativas e a sua variante CMYK, usada maioritariamente em impressoras, usa também o preto, pois esta cor não é obtida através da combinação das outras três [47]. Outros modelos como o HSV e o HLS foram criados para reproduzir um processo de mistura de cores mais intuitivo, como o usado na pintura. São particularmente úteis para a seleção de cores em programas de arte e representam a cor em componentes distintos como: o tom (H), a saturação (S) e o brilho (L) para o HLS e o tom (H), a saturação (S) e o valor (V) para o HSV [13, 47]. Por último, o modelo YUV, foi uma das primeiras representações de cor criadas para transmissão de vídeo. Este possui três componentes: um canal Y que representa a luminância e dois outros canais: U e V, para representar a cor. O mapeamento entre RGB e YUV é dado pela fórmula 2.3:

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ U &= 0.492111(B - Y) \\ V &= 0.877283(R - Y) \end{aligned} \quad (2.3)$$

Todas as conversões entre os modelos de cor usados neste trabalho são feitas com recurso à biblioteca *OpenCV* para processamento de imagem [27].



(a) Cores aditivas.



(b) Cores subtrativas.

Figura 2.2: Cores primárias e cores secundárias formadas a partir da mistura das cores aditivas (a) e subtrativas (b) [47].

### 2.1.2 Descritores de Características da Imagem

Para analisar um conjunto de imagens é preciso estabelecer um valor com o qual as identificar e comparar. Os descritores são conjuntos de valores que descrevem características das imagens, podendo estar relacionados com a sua cor, textura, formas, contornos, entre outros. Devem ser únicos, simples (processamento rápido) e conter o necessário para representar a imagem sem perda de informação [18].

#### 2.1.2.1 Descritores de Cor

O histograma é um descritor que identifica a imagem segundo a distribuição do seu número de *pixels*. O histograma mais comum é o RGB que consiste na distribuição do número de *pixels* por cada canal de cor [39]. No entanto, os histogramas não têm em conta a informação da localização dos *pixels* da imagem e, conseqüentemente, duas imagens diferentes podem obter a mesma distribuição de cor. No caso do descritor *Color Coherent Vector* (CCV), os *pixels* são divididos em coerentes, caso estejam ligados a uma zona da mesma cor, e incoerentes caso estejam isolados [18], o que permite diferenciar as duas imagens.

Uma distribuição de probabilidade pode ser caracterizada pelos seus momentos. Se interpretarmos a distribuição da cor como uma distribuição de probabilidade então esta também pode ser descrita pelos seus momentos. Os três momentos principais são a média dos *pixels* da imagem, o desvio padrão e a assimetria de cada canal de cor [45]. Estes são relativamente simples e eficientes na representação da imagem.

Existem ainda descritores calculados a partir do algoritmo *Scale Invariant Feature Transform* (SIFT) que são invariantes a translações, rotações e à escala. Dependendo do tipo de descritor escolhido ou o modelo de cor aplicado, o descritor pode ter propriedades invariantes a certos fatores como o ângulo, a dimensão e a iluminação da imagem [39].

#### 2.1.2.2 Contornos

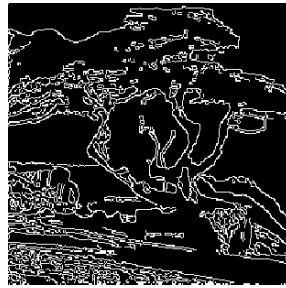
O contorno de um objeto é suficiente para o representar de modo a que este possa ser facilmente identificado numa imagem. Estes ocorrem onde há uma rápida uma variação de intensidade na imagem [47] e geralmente a sua deteção é feita através da aplicação de filtros de convolução, mas existem também métodos mais complexos.

Os métodos de gradiente como o *Roberts*, *Prewitt* e *Sobel* baseiam os seus filtros na primeira derivada da curva de intensidade, obtendo para cada *pixel* um valor do gradiente local. Caso esse valor seja superior a um *threshold* definido o *pixel* será marcado como contorno [31]. Existem também métodos como o *Laplace* que recorrem à segunda derivada nas suas máscaras para procurar obter menos *pixels* de contorno do que os métodos de gradiente [47]. Antes da aplicação destes operadores é comum aplicar um filtro para redução de ruído como, por exemplo o gaussiano, pois estes métodos podem ser bastante sensíveis à existência de ruído na imagem.

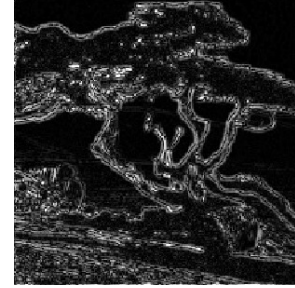
Um método mais avançado de detecção é o algoritmo de *Canny*, que pode ser definido em quatro passos. Inicialmente é aplicado um filtro gaussiano para remover o ruído. De seguida são aplicados os filtros de *Sobel* para obter as primeiras derivadas nas direções vertical e horizontal, a partir das quais se calcula o valor do gradiente da imagem e a sua direção para cada *pixel*. No terceiro passo são removidos os *pixels* que não são considerados como contorno, verificando quais são um máximo local. Por fim são utilizados dois *thresholds* (um superior e outro inferior) para decidir quais os *pixels* marcados anteriormente são de facto parte do contorno da imagem. Caso o gradiente do *pixel* seja maior que o valor superior é aceite como contorno, caso seja inferior ao valor mínimo é descartado e, caso fique entre os dois valores definidos, é aceite se estiver ligado a um *pixel* já aceite como parte do contorno [26].



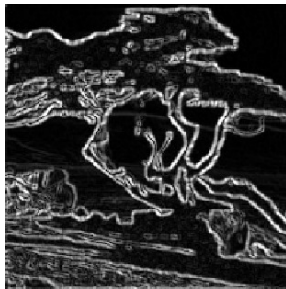
(a) Imagem original.



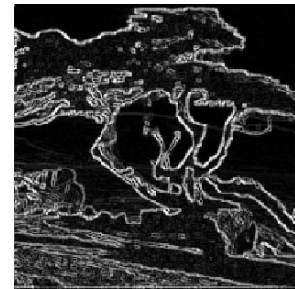
(b) Canny.



(c) Laplace.



(d) Sobel.



(e) Prewitt.

Figura 2.3: Comparação entre diferentes filtros para detecção de contornos [2].

### 2.1.2.3 Forma

Apesar dos contornos conseguirem descrever os objetos, por vezes pode ser útil detetar linhas ou determinada forma na imagem. A transformada de *Hough* é um método que permite encontrar linhas, círculos, elipses ou outras formas simples numa imagem. A sua versão original é uma forma rápida de procurar linhas retas numa imagem binária, onde cada ponto pertencente ao contorno vota em todas as linhas possíveis que possam passar por si [47].

Considerando um sistema de coordenadas polares, uma linha reta é representada por  $r = x \cos \theta + y \sin \theta$ . Para cada ponto  $(x_i, y_i)$  da imagem podemos definir o conjunto de potenciais linhas que passam por ele como  $r_i \theta = x_i \cos \theta + y_i \sin \theta$  (Figura 2.4a). Assim, cada

par  $(r_i, \theta_i)$  representa uma linha que passa no ponto  $(x_i, y_i)$ .

Estes conjuntos de potenciais linhas formam sinusoides no plano  $(r, \theta)$ , o plano acumulador (Figura 2.4b). Quando as sinusoides se intersectam os pontos que as originaram pertencem a uma mesma linha. Assim, quantos mais pontos pertencerem a uma linha, mais esta é relevante para a deteção. Ao definir um *threshold* mínimo de pontos pertencentes a uma linha, podemos considerar que todas as linhas acima desse valor são relevantes e assim selecionar quais são detetadas [47].

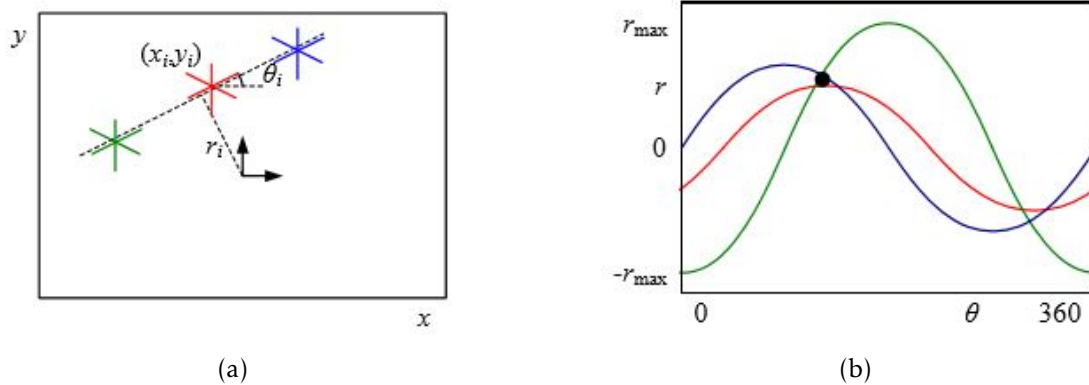


Figura 2.4: Transformada de Hough: (a) Conjunto de potenciais linhas que passam por cada ponto  $(x_i, y_i)$ ; (b) Intersecção das sinusoides formadas pelos conjuntos de linhas de cada ponto. A sua intersecção dá origem a uma linha detetada.

A versão da transformada dedicada aos círculos funciona de forma semelhante à original, mas procura aumentar a eficiência através de um método chamado *Gradiente de Hough*, no qual é usado um acumulador bidimensional para registar as interseções de cada ponto da imagem. Numa primeira fase é feita a deteção dos possíveis centros de círculos na imagem. Inicialmente é aplicado à imagem o algoritmo *Canny* a partir do qual se obtém uma imagem binária com os contornos, sendo que para cada ponto não nulo dessa imagem é calculado o gradiente a partir das derivadas dos operadores de *Sobel*  $x$  e  $y$ . De seguida é necessário incrementar o acumulador nos valores referentes aos pontos que pertencem às linhas dos gradientes cuja distância está dentro de um certo intervalo. A partir dos pontos registados no acumulador, são escolhidos como possíveis centros aqueles cujo valor for superior a um certo *threshold* e ao valor dos seus vizinhos imediatos. Esses possíveis centros são ordenados de forma decrescente dos seus valores no acumulador, de modo a que os centros com mais votos surjam primeiro.

Na segunda fase é escolhido o melhor raio para cada um dos centros selecionados. Para cada possível centro, todos os *pixels* não nulos são ordenados segundo a sua distância ao centro. Partindo das distâncias mais pequenas até ao raio máximo, escolhe-se o raio que é apoiado pelo maior número de *pixels*.

O centro selecionado é mantido caso tenha apoio suficiente dos *pixels* não nulos e esteja a distância suficiente de outro centro já escolhido [5].



### 2.1.3 Classificação

Aprendizagem automática é uma área da Inteligência Artificial que procura construir sistemas capazes de utilizar dados para melhorar o seu desempenho numa determinada tarefa. É um conceito que abrange muitas áreas de aplicação sendo que para cada problema diferente é preciso uma abordagem distinta. Permite resolver os problemas sem programar a sua solução diretamente, sendo apenas necessário fornecer um conjunto de dados ao algoritmo para que este consiga, com ou sem supervisão, prever o resultado corretamente. Caso seja usado um conjunto de dados que contenha os valores das respostas corretas para que o algoritmo possa extrapolar os novos resultados, estamos perante aprendizagem supervisionada. Se o conjunto de dados não estiver etiquetado com essas informações trata-se de aprendizagem não supervisionada [1].

No caso do processamento de imagem as aplicações mais comuns são o reconhecimento e a localização de objetos numa imagem. Estas relacionam-se diretamente com os problemas de classificação, onde o objetivo é prever a que categoria pertence um determinado exemplo, e de *clustering*, onde se procura agrupar os dados naturalmente [5]. A análise colorimétrica a realizar neste projeto é um problema de classificação com múltiplas classes.

#### 2.1.3.1 Processamento dos Dados

Antes de proceder a qualquer implementação de algoritmo é necessário fazer um tratamento dos dados para que estes permitam obter o melhor resultado possível. Neste contexto, os dados utilizados serão provenientes de descritores de imagem como os referidos na secção 2.1.2.

O processamento segue os seguintes passos [5]:

1. **Recolher e etiquetar os dados** - Os dados recolhidos devem ser relevantes para o problema e etiquetados consoante o resultado que se procura atingir.
2. **Extraír descritores de características** - Escolher os descritores a extrair consoante os objetivos do problema. Estes devem ser relevantes e apresentar alguma invariância no objeto a identificar. São depois convertidos para um vetor de características (*features*). Para obter as *features* relevantes pode ser preciso um processamento adicional para transformar o conjunto de dados originais num conjunto mais útil para o problema. O algoritmo geralmente aplicado para estas situações é a Análise de Componentes Principais (PCA) que permite reduzir a dimensão do conjunto de dados com perda mínima de informação [1].

Alguns algoritmos podem também ser negativamente influenciados pela diferença de escalas entre os valores como, por exemplo, aqueles que dependem de medidas de distância [5]. Assim, caso os descritores extraídos tenham valores de escalas muito diferentes é necessário fazer a standardização dos dados para que estes tenham

todos a mesma ordem de grandeza. Tipicamente recorre-se à fórmula 2.4, onde  $\mu$  representa a média do conjunto de dados e  $\sigma$  o desvio padrão [32].

$$x' = \frac{x - \mu}{\sigma} \quad (2.4)$$

No entanto, o conjunto de dados pode ter *outliers*, isto é, valores invulgarmente altos ou baixos em relação aos restantes dados [50]. Estes podem influenciar negativamente a média e a variância do conjunto, o que pode levar a piores resultados utilizando o método anterior. Nesse caso devemos recorrer a métodos estatísticos que lidem melhor com os *outliers* como, por exemplo, a fórmula 2.5 onde  $\tilde{x}$  representa a mediana do conjunto de dados e  $Q_3 - Q_1$  refere-se à amplitude interquartil.

$$x' = \frac{x - \tilde{x}}{Q_3 - Q_1} \quad (2.5)$$

A amplitude interquartil é a diferença entre o primeiro quartil, isto é, o valor a antes do qual estão 25% dos valores menos elevados e o terceiro quartil, que é o valor a partir do qual estão 25% dos valores mais elevados do conjunto. Deste modo, não é influenciada por valores extremos [50].

3. **Dividir os dados** - Os dados processados anteriormente são divididos em três conjuntos diferentes: o de treino é usado para treinar o algoritmo, o de validação para validar os resultados do algoritmo ou prever parâmetros e o conjunto de teste para estimar a capacidade de previsão do classificador para dados fora do conjunto de treino. Antes de dividir os dados pode ser feita uma mistura aleatória para eliminar qualquer ordem prévia que possa existir nos mesmos.

A divisão dos dados geralmente envolve um processo chamado *K-Fold Cross-Validation*, que permite obter uma melhor estimativa da precisão com que o nosso modelo irá prever novos dados (diferentes do conjunto de treino). Os dados são divididos em  $K$  subconjuntos (*folds*), sendo o modelo treinado com  $K - 1$  *folds* e validado com a *fold* restante. Este processo é repetido para que todas as *folds* possam ser usadas como validação e no final é feita a média dos erros de validação que nos dá a estimativa do verdadeiro erro do modelo [1].

A validação é um processo importante, pois caso o classificador seja otimizado em excesso pode vir a adaptar-se demasiado ao conjunto de dados. Como consequência este perde a capacidade de prever a classe de amostras que não tenha observado durante o processo de treino, sendo esta situação descrita como *overfitting* [1].

### 2.1.3.2 Algoritmos para Classificação

Com os dados preparados é preciso escolher o classificador a utilizar, sendo que cada um tem as suas vantagens. Caso a classificação envolva mais que duas classes, pode ser necessário fazer algumas alterações ao utilizar determinados algoritmos que estejam preparados apenas para classificações binárias [1]. De seguida são descritos alguns dos algoritmos mais usados para classificação com imagens e ainda outros tipos de algoritmos, como o *clustering* hierárquico, mencionados nas secções seguintes do documento.

***K-Nearest Neighbours (KNN)*** Um classificador simples que guarda o conjunto de treino com as suas etiquetas e classifica um novo ponto com a etiqueta da maioria dos  $K$  pontos no conjunto de treino que estão mais perto deste novo ponto (os seus vizinhos). A função usada para calcular a distância pode ser definida consoante o tipo de dados [1]. É considerado eficiente, mas é algo lento e requer muita memória [5].

***Support Vector Machines (SVM)*** Procura encontrar um hiperplano ótimo num espaço de  $N$  dimensões, onde  $N$  é o número de *features*, que separe os pontos em classes distintas. O objetivo é encontrar o hiperplano que separe corretamente as classes maximizando a margem, isto é, a distância entre o hiperplano e os pontos dessas duas classes [1]. Este processo é usado maioritariamente para classificação, mas também consegue fazer regressão, atingindo muito bons resultados com conjuntos de dados limitados [5]. Uma das suas variantes, o *Least Squares Support Vector Machines* (LS-SVM) encontra o hiperplano ótimo através de um conjunto de equações lineares, enquanto que o SVM resolve um problema de programação quadrática. Deste modo, o LS-SVM consegue obter um custo de processamento mais baixo que o SVM normal [46].

Caso os dados analisados não sejam linearmente separáveis, isto é, separáveis com uma linha reta, é necessário transformar o conjunto de dados para uma dimensão superior que permita a sua separação linear [1]. Os *kernels* são as funções que permitem transformar o conjunto de dados original para a dimensão desejada para a sua separação. Existem vários tipos de *kernel* (figura 2.5). O mais simples, o linear, obtém um hiperplano linear, enquanto outros como o RBF e o polinomial permitem obter hiperplanos não lineares [11]. No contexto desta dissertação, os dados analisados não são linearmente separáveis. Assim, as implementações do SVM utilizadas recorrem ao *kernel* RBF.

Para estes modelos existem dois parâmetros que influenciam significativamente o resultado da classificação e devem ser selecionados cuidadosamente [11]. O parâmetro  $C$  é um parâmetro de regularização que penaliza a classificação incorreta dos exemplos. Um  $C$  pequeno leva a uma maior tolerância de classificações incorretas, o que se traduz numa margem de decisão maior, podendo levar a *underfitting*. Por outro lado, um  $C$  elevado leva a um modelo muito restrito e ao conseqüente *overfitting* [1]. Em relação ao segundo parâmetro  $\gamma$ , quanto maior o seu valor mais o hiperplano se irá ajustar ao modelo de

dados, o que irá levar ao *overfitting*. No entanto, caso seja muito pequeno o modelo perde toda a capacidade de se adaptar aos dados, comportando-se como um modelo linear [32].

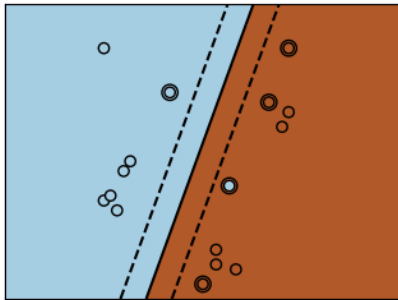
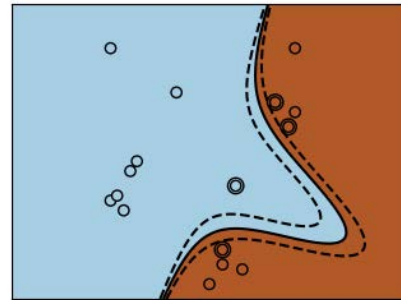
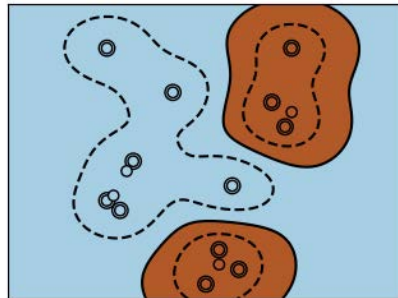
(a) *Kernel Linear*(b) *Kernel Polinomial*(c) *Kernel RBF*

Figura 2.5: Exemplo da classificação de SVM com três *kernels* diferentes: (a) linear, (b) polinomial e (c) RBF. As linhas tracejadas mostram os limites das margens enquanto que as linhas normais indicam o hiperplano que separa as duas classes [32].

**Random Trees** Um conjunto de múltiplas árvores de decisão, cada uma construída até uma certa profundidade máxima. Uma árvore de decisão é composta por nós de decisão e folhas terminais. Cada nó de decisão tem uma função com um resultado definido para cada ramo. Dado um determinado valor de entrada, o nó é testado e o ramo a seguir é escolhido consoante o resultado da função. Este processo começa na raiz e é repetido recursivamente até que se atinja um nó folha, cujo valor será o resultado final [1]. Durante o treino, cada nó de cada árvore determina a sua divisão a partir de um conjunto aleatório de *features*, o que ajuda a assegurar que cada árvore final se torna um decisor estatisticamente independente. Ao executar, cada uma das árvores tem um voto de decisão. É considerado um algoritmo muito eficiente sendo também capaz de fazer regressão [5].

### 2.1.3.3 Outros Algoritmos

**Hierarchical Cluster Analysis (HCA)** Este algoritmo de *clustering* procura agrupar os dados de modo a que os membros do grupo sejam mais semelhantes entre si do que outros pontos de outros grupos. Assim, é definida uma medida de semelhança, nomeadamente uma distância entre os pontos. Pode ser de aglomeração, onde se começa com N grupos, cada um com um elemento do conjunto de dados, e se vai unindo os conjuntos mais próximos até haver apenas um, ou divisão, onde se divide um conjunto maior em conjuntos menores [1]. O resultado do algoritmo pode ser representado por um dendograma que permite visualizar bem os dados, estabelecendo uma estrutura hierárquica que pode ser útil para a análise.

### 2.1.3.4 Avaliação

Para além da estimativa a partir do erro de validação já referida anteriormente, existem outras formas de avaliar os classificadores, baseadas na sua exatidão. Para isso precisamos primeiro definir as quatro possibilidades possíveis para a classificação [1]:

- Verdadeiro Positivo ( $tp$ ) - O exemplo pertence à classe para a qual foi previsto.
- Falso Positivo ( $fp$ ) - O exemplo não pertence à classe para a qual foi previsto.
- Verdadeiro Negativo ( $tn$ ) - O exemplo não pertence à classe para a qual não foi previsto.
- Falso Negativo ( $fn$ ) - O exemplo pertence à classe para a qual não foi previsto.

A partir destas possibilidades podemos calcular várias métricas para avaliar o classificador. Apesar de estas métricas serem aplicadas em problemas de classificação binária podem ser generalizadas para utilização em casos que lidem com múltiplas classes. Existem duas opções para esta generalização: a micro, que favorece classes com mais amostras e a macro, que trata todas as classes de igual modo [43]. No contexto deste trabalho existem classes com grande diferença de amostras entre si. Como se pretende que todas tenham a mesma importância nos resultados das métricas de seleção dos algoritmos será aplicada a macro.

- **Exatidão (*Accuracy*)** - fração das classificações feitas corretamente.

$$Accuracy = \frac{tp + tn}{tp + fn + fp + tn} \quad [1] \quad (2.6)$$

Para problemas multi-classe aplicamos a seguinte fórmula, onde  $l$  corresponde ao número total de exemplos:

$$Accuracy_{average} = \frac{\sum_{i=1}^l \frac{tp_i + tn_i}{tp_i + fn_i + fp_i + tn_i}}{l} \quad [43] \quad (2.7)$$

- **Precisão (*Precision*)** - proporção de classificações positivas corretas [1]. No contexto de diagnóstico de uma doença, por exemplo, num classificador com precisão igual a um, todos os indivíduos classificados como doentes têm a doença, mas podem existir indivíduos doentes não diagnosticados.

$$Precision = \frac{tp}{tp + fp} \quad [1] \quad (2.8)$$

$$Precision_{Macro} = \frac{\sum_{i=1}^l \frac{tp_i}{tp_i + fp_i}}{l} \quad [43] \quad (2.9)$$

- **Sensibilidade (*Recall*)** - permite identificar a proporção de valores positivos que foi classificada corretamente [1]. Se classificador tem a sensibilidade igual a um, todos os indivíduos doentes foram diagnosticados, mas pode haver pessoas saudáveis diagnosticadas incorretamente como portadores da doença.

$$Recall = \frac{tp}{tp + fn} \quad [1] \quad (2.10)$$

$$Recall_{Macro} = \frac{\sum_{i=1}^l \frac{tp_i}{tp_i + fn_i}}{l} \quad [43] \quad (2.11)$$

- **$F_\beta$ Score** - Medida de exatidão do classificador baseada na *Precision* e no *Recall* [43]. O parâmetro  $\beta$  permite alterar o impacto da precisão ou da sensibilidade consoante as especificidades do nosso problema, sendo que quando  $\beta = 1$  este corresponde à média harmónica destes valores. No decorrer deste trabalho foi usado o *FScore* com  $\beta = 1$ .

$$F_\beta Score = \frac{(\beta^2 + 1)tp}{(\beta^2 + 1)tp + \beta^2 fn + fp} \quad [43] \quad (2.12)$$

$$F_\beta Score_{Macro} = \frac{(\beta^2 + 1)Precision_{Macro}Recall_{Macro}}{\beta^2 Precision_{Macro} + Recall_{Macro}} \quad [43] \quad (2.13)$$

## TRABALHO RELACIONADO

Neste capítulo é feito um estudo de trabalhos de investigação que abordam conceitos e técnicas relacionadas com os objetivos que se procura atingir com a presente dissertação. Os tópicos serão abordados em duas secções. Na primeira secção são apresentados sistemas que realizam análise colorimétrica de biossensores e na segunda são descritos métodos de calibração de cor aplicados em diversas áreas científicas.

### 3.1 Análise Colorimétrica

Existem vários projetos desenvolvidos que assentam na análise colorimétrica de biossensores. Estes sistemas destinam-se à deteção dos níveis de substâncias químicas como a glucose, o pH e o ácido láctico em compostos como o suor e a urina. Permitem a realização de uma análise qualitativa e por vezes quantitativa destas substâncias.

Nesta secção vão ser apresentados alguns trabalhos que envolvem processos de análise colorimétrica aplicados a biossensores. Estes envolvem diferentes processos de extração de cor e tratamento de imagem bem como a classificação de valores de concentrações através de comparações com curvas de calibração e da aplicação de diversos algoritmos de aprendizagem automática.

Após uma breve descrição dos diferentes processos será introduzida a abordagem aplicada na versão anterior da aplicação e por fim será feita uma discussão das diversas vantagens e problemas que cada um deles apresenta, com o objetivo de perceber que partes serão melhor aplicáveis ao contexto do trabalho a desenvolver na dissertação.

#### 3.1.1 Observação de Curvas de Calibração

As curvas de calibração são instrumentos utilizados para análises químicas. São realizados vários ensaios com amostras de concentrações conhecidas da substância que se

pretende analisar e a partir desses resultados obtêm-se gráficos que permitem extrapolar as concentrações para amostras desconhecidas [33].

No projeto de Koh et al. [17] foi desenvolvida uma aplicação para leitura de um biosensor autocolante para analisar o suor da pele através da captura de imagens. O sensor mede a taxa de suor libertado e analisa-o para descobrir o valor do pH e as concentrações de cloreto, ácido láctico e glucose. Recorre-se à tecnologia *Near Field Communication* (NFC) para que quando o utilizador aproxime o *smartphone* do sensor a aplicação inicie a captura de imagens do mesmo. A informação RGB da cor da região detetada é então extraída e mostrada no ecrã da aplicação. Após uma calibração para correção da cor, o valor da cor em RGB% é convertido para a concentração correspondente através da relação com as curvas de calibração pré-calculadas mostradas na figura 3.1.

O sensor foi testado em exercício físico de intensidade média e atividade desportiva de alta competição, sendo que a análise apresentou resultados favoráveis sendo possível detetar mudanças de até 0,5 unidades de pH e de 0,3 a 0,1 mM nas concentrações da glucose e dos restantes compostos.

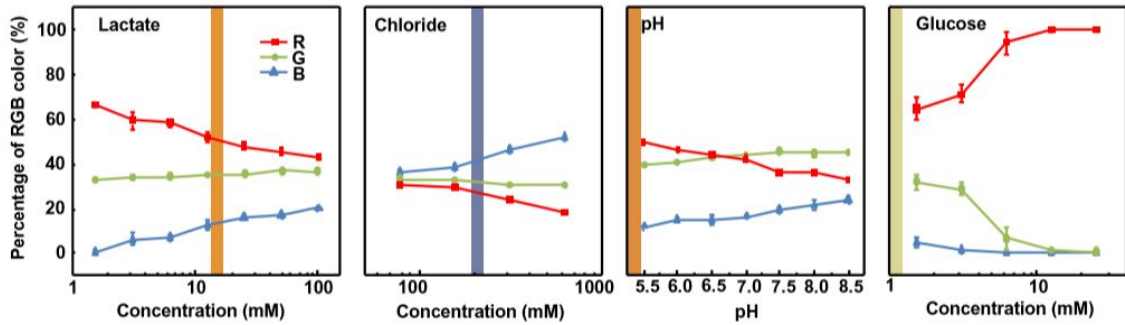


Figura 3.1: Curvas de calibração que relacionam a percentagem de RGB com a concentração da substância analisada [17]

Yetisen et al. [52] desenvolveu uma aplicação para leitura de testes comerciais de urina, sendo possível medir o valor do pH, proteína e glucose a partir da análise da imagem dos mesmos. Para analisar uma amostra é preciso primeiro fazer um passo de calibração da aplicação. Este processo cria uma curva de calibração para um certo tipo de teste a uma determinada iluminação e distância, que é depois guardada. Caso se pretenda fazer um teste com outras condições de iluminação é necessário voltar a calibrar.

O processo de construção da curva é semelhante ao utilizado em [41], onde se verificou a existência de correlações entre os valores de pH e os valores da cor no espaço CIE XYZ. Para construir esta curva, a aplicação localiza as cores de referência e transforma a informação obtida através da câmara para valores RGB não lineares. Esses valores são depois convertidos para valores RGB lineares através das equações seguintes:

$$R_l = \left( \frac{0.055 + R_{notL}}{1.055} \right)^{2.4} \quad G_l = \left( \frac{0.055 + G_{notL}}{1.055} \right)^{2.4} \quad B_l = \left( \frac{0.055 + B_{notL}}{1.055} \right)^{2.4} \quad (3.1)$$



Os valores RGB lineares de cada ponto de calibração são então convertidos para o espaço CIE XYZ e finalmente para as coordenadas  $x$  e  $y$  do diagrama cromático CIE, através de processos descritos na secção 2.1.1.1. Após calcular estes valores para o número diferente de pontos de calibração desejados e introduzir também as respectivas concentrações que estes representam, a aplicação guarda todos os valores numa base de dados interna como a curva de calibração para aquela configuração.

A fase da análise colorimétrica consiste na comparação dos valores medidos com a curva de calibração que corresponde às mesmas condições da medição. A partir da imagem capturada pelo utilizador, é repetido o processamento feito na construção da curva até à obtenção dos valores do espaço CIE XYZ. Aí, para cada ponto da curva de calibração  $j$ , calcula-se a distância euclidiana entre o valor medido e  $j$ .

O algoritmo guarda as duas distâncias mais curtas ao ponto da amostra ( $d_{ks}$  e  $d_{kss}$ ), bem como os seus valores ( $x, y$ ) e as suas concentrações, calculando também a distância entre elas ( $d_C$ ). São calculadas também a distância euclidiana entre os dois pontos mais próximos na curva e o ponto medido ( $d_{xy}$ ) e a distância mais curta entre o ponto medido e a linha entre os dois pontos de calibração ( $d_{sd}$ ), calculada em 3.2. O valor correspondente ao máximo entre essas 3 distâncias ( $d_{ks}, d_{kss}, d_{sd}$ ) é então guardado como  $d_m$ .

$$d_{sd} = \frac{|(x_{kss} - x_{ks})(y_{ks} - y_m) - (x_{kss} - x_m)(y_{ks} - y_{kss})|}{\sqrt{(x_{kss} - x_{ks})^2 + (y_{ks} - y_{kss})^2}} \quad (3.2)$$

A variação ( $v \pm$ ) da concentração é dada pela fórmula 3.3 e o valor da concentração ( $C_m$ ) é obtido calculando a distância de um dos pontos de calibração ao ponto na linha onde  $d_{sd}$  é medida (Equações 3.4).

$$v \pm \frac{d_m d_C}{d_{xy}} \quad (3.3)$$

$$C_m = \frac{d_p d_C}{d_{xy}} \quad (3.4)$$

$$d_p = \sqrt{d_{ks}^2 - d_{sd}^2}$$

O sistema foi testado com tiras de testes mergulhadas em urina artificial para obter mais controlo sobre os valores que se pretendia confirmar e a calibração foi feita com recurso ao quadro de referência incluído na embalagem dos testes sob luz fluorescente, sendo possível observar resultados precisos na medição dos vários compostos.

### 3.1.2 Abordagens com recurso a Aprendizagem Automática

O trabalho desenvolvido por Mutlu et al. [24] consiste na aplicação de algoritmos de aprendizagem automática à análise colorimétrica de tiras de pH através do *smartphone*. Para classificar os valores obtidos são utilizados dois classificadores: o SVM e uma das suas variantes, o LS-SVM, descritos na secção 2.1.3.2.

Para treinar os classificadores a utilizar é preciso primeiro obter um conjunto de dados. De modo a construir esse conjunto foram captadas imagens em duas experiências

distintas, sendo que em ambas foram fotografadas tiras com valores de pH de zero a 14, experimentando com seis orientações diferentes para obter iluminação variável, mas mantendo sempre a mesma distância. Na primeira experiência foi utilizado um adereço para bloquear a iluminação exterior, sendo o *flash* da câmara a única fonte de luz disponível enquanto que na segunda experiência sem adereço foram observadas três iluminações diferentes (solar, fluorescente e halogéneo). Foram ainda testados pares de combinações possíveis destas iluminações, mas apenas com tiras de pH entre quatro e nove. As imagens foram capturadas com definições de câmara constantes e guardadas nos formatos JPEG e RAW, sendo obtidas para a primeira experiência 450 imagens e para a segunda 270. Estas têm depois de ser cortadas e posicionadas numa posição vertical, prontas para a extração de cor.

- **Extração da cor** - Existem tiras com quatro zonas de teste e em três formatos de imagem diferentes (JPEG, RAW e RAW corrected). Inicialmente é feita a extração dos valores médios de cada canal RGB e criada uma matriz  $4 \times 3$  de *features*  $X_n(i, j)$ , onde  $i$  é a zona de teste e  $j$  a cor RGB para cada imagem  $n$ . Cada uma dessas matrizes é depois mapeada para um vetor de dimensão  $12 \times 1$   $X_n$ , devidamente etiquetado para obtermos um conjunto de treino com 15 classes diferentes (valores pH de zero a 14).
- **Classificação dos dados** - Para classificar os dados extraídos recorreu-se a dois classificadores de aprendizagem automática, o LS-SVM e o SVM. Após passar as *features* extraídas ( $X_n$ ) para o classificador é possível estimar quão precisamente este distingue as tiras de pH de um conjunto de dados mais generalizado aplicando o método de *K-fold Cross Validation* descrito na secção 2.1.3.1, onde  $K=10$ .

De modo a avaliar a prestação dos classificadores foram calculadas três métricas: precisão, *recall* e *specificity*. O classificador LS-SVM obteve melhores resultados que o SVM, sendo que com isolamento da luz exterior obteve-se uma precisão de 100% e de aproximadamente 100% nas experiências sem o aparelho. Nos casos em que existiam duas fontes de luz presentes obteve-se resultados acima de 80%. Esta diferença é esperada porque foram usadas para treino imagens com uma só iluminação. Constataram também que o formato das imagens não condicionou o resultado dos testes, pelo que se pode recorrer a imagens JPEG que exigem menos processamento que as RAW.

Jia et al. [14] propôs um método de calibração de cor para biossensores que detetam a glucose na urina utilizando o *smartphone*. Este método de calibração encontra-se descrito na secção 3.2, tendo sido também realizadas experiências com vista a classificar amostras para demonstrar a qualidade da calibração, daí a sua descrição nesta secção. Estas experiências foram realizadas com um dos métodos de calibração propostos, o dos quatro pontos. Para agrupar as amostras foi aplicado à base de dados o algoritmo de *clustering* hierárquico HCA, descrito na secção 2.1.3.2.

Para estabelecer a base de dados de amostras necessária para a comparação, foram fotografadas amostras com seis concentrações diferentes e um controlo antes e depois da reação química que provoca a mudança da cor. Estas experiências foram repetidas cinco vezes por concentração, com uma iluminação fluorescente e uma distância fixa. As cores das imagens foram depois corrigidas e foi calculada a diferença dos valores RGB entre as imagens antes e depois da reação. Para testar o processo de calibração foram fotografados sensores com as mesmas concentrações anteriores, mas desta vez com condições de iluminação aleatórias. Essas imagens são depois corrigidas e a diferença entre o antes e depois da reação é inserida na base de dados. Por último, os dados foram analisados usando o algoritmo HCA, verificando-se que os novos valores testados foram agrupados nos locais corretos (Figura 3.2), o que demonstra o sucesso da calibração.

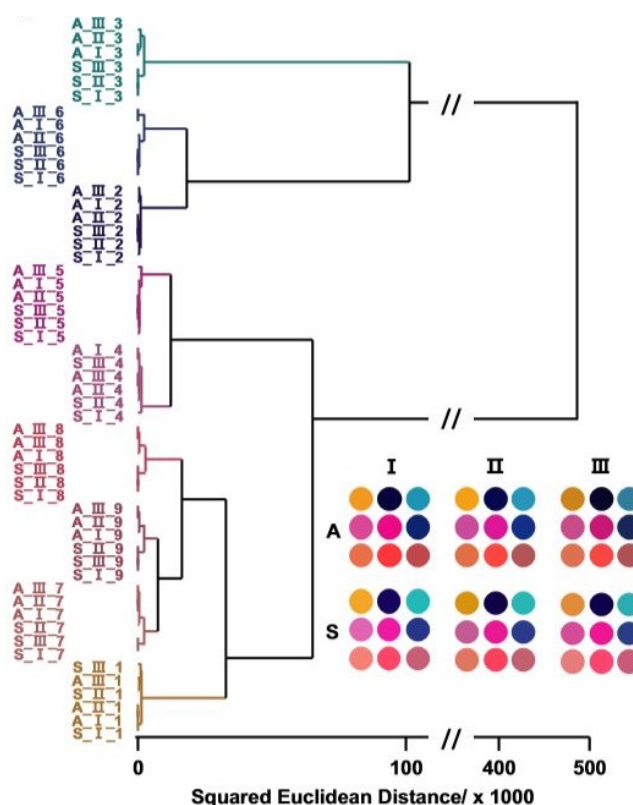


Figura 3.2: Análise com HCA das cores do sensor após calibração para seis experiências. Cada experiência sob o efeito da luz solar exterior (I), luz fluorescente interior (II), luz interior fraca (III) para dois *smartphones* diferentes: *iPhone 4* (A) e *Samsung Galaxy S II* (B). [14]

Verificaram ainda a existência de uma boa relação entre concentrações de glucose logarítmicas e os resultados do sensor, o que pode ser útil para uma análise quantitativa precisa da glucose. Foram também testadas amostras de urina de diferentes concentrações das calibradas. Estas foram classificadas corretamente, o que indica que esta técnica pode ser aplicada para realizar análises colorimétricas desta natureza.

### 3.1.3 Versão Anterior da Aplicação

Na versão anterior da aplicação desenvolvida por Rodrigues et al. [38] realizou-se a análise colorimétrica de um biossensor para concentrações de glucose com recurso a classificação através de algoritmos de aprendizagem automática. Inicialmente é feita a deteção das zonas de interesse no sensor através do processo descrito na secção 6.1. Após detetar a região a analisar é feita a extração da cor calculando a média dos seus *pixels* no formato RGB. A zona é reduzida a 80% do raio do círculo detetado para que os *pixels* da parte preta do sensor não influenciem a cor extraída caso intersectem essa zona. São ainda extraídas as cores correspondentes às zonas de referência usadas na seguinte fase de calibração descrita em 3.2.3.

Para analisar os dados foram criados dois modelos para classificadores de aprendizagem automática. O primeiro destina-se a analisar dados do utilizador em estado de jejum e tem por base o classificador SVM no espaço de cor HLS. O segundo modelo foca-se em dados adquiridos duas horas após a refeição e utiliza o classificador KNN e o espaço de cor RGB, descritos na secções 2.1.1 e 2.1.3. Os classificadores foram treinados com um conjunto de dados formado por sensores com diferentes concentrações fornecidos pelo CENIMAT, etiquetados de acordo com a sua concentração baseada em valores de referência para a diabetes detalhados na tabela 3.1.

	Hipoglicemia	Normal ou Pré-diabetes	Diabetes
Em jejum	0-70 mg/dl	70-126 mg/dl	>126 mg/dl
2 horas após a refeição	0-70 mg/dl	70-200 mg/dl	>200 mg/dl

Tabela 3.1: Classes de concentração de glucose usadas nos modelos

Após a conversão para os modelos de cor escolhidos e antes da aplicação dos classificadores é necessário fazer o pré-processamento dos dados. Este processo consiste numa operação de normalização através da fórmula 3.5 seguida do processo de PCA descrito na secção 2.1.3.1. Este processo permite uma melhor visualização dos dados em 2D, como pode ser observado na figura 3.3.

$$x' = \frac{x - \bar{x}}{\sigma_x^2} \quad (3.5)$$

Foram realizados testes a par com a análise, tendo sido testados vários classificadores (SVM, KNN e *Random Trees*) e modelos de cor (HSV, XYZ, YUV, HLS, RGB, rácio entre R e G) até se obter os escolhidos como melhor enquadrados no contexto do problema. Para testar o processo de análise foram usadas várias imagens com diferentes concentrações de glucose e iluminação (solar, fluorescente e incandescente), capturadas a partir de três *smartphones* diferentes. Os modelos de classificação foram avaliados com *K-Fold Cross Validation* de  $K = 10$  e as métricas de precisão, *recall* e *F1-score* foram as escolhidas para avaliar o desempenho dos algoritmos.

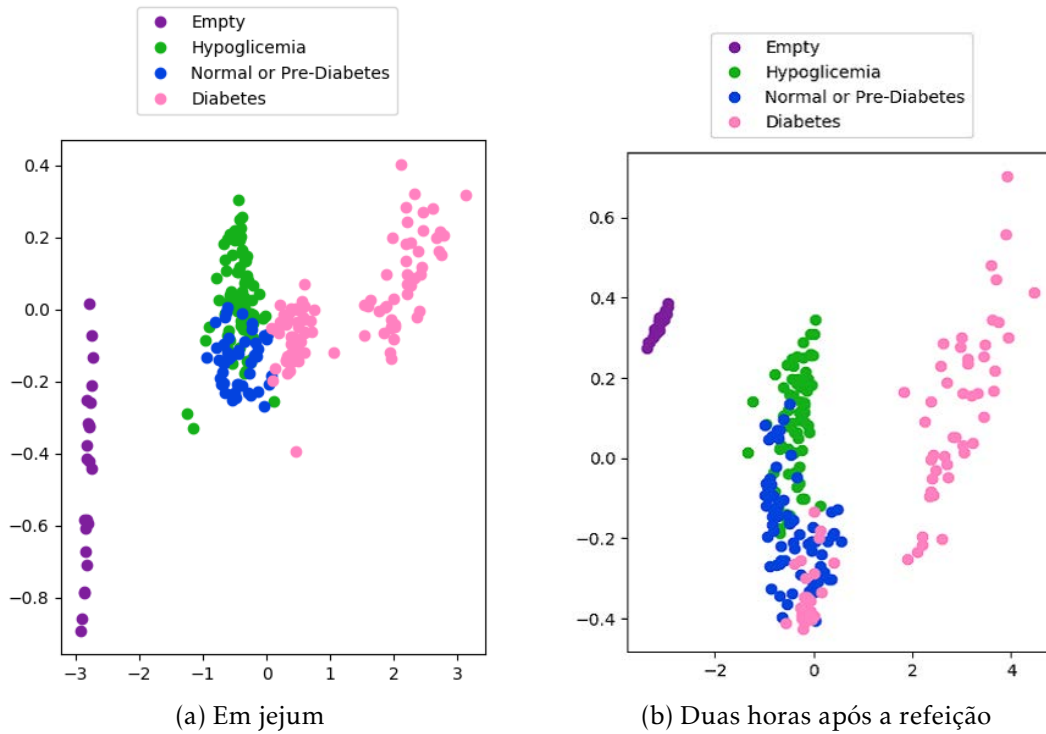


Figura 3.3: Projeção PCA dos conjuntos de dados [38]

Os testes revelaram que seria necessária uma maior amostra para o conjunto de dados. Também seria necessário um melhor processo de calibração da cor, pois em certas câmaras e fracas iluminações os resultados foram influenciados negativamente. Concluiu-se também que os modelos escolhidos têm uma capacidade para o reconhecimento preciso dos biossensores vazios, com baixas concentrações (hipoglicemia) e com altas concentrações (hiperglicemia). No entanto, os resultados obtidos com as concentrações intermédias não são tão precisos.

### 3.1.4 Análise Crítica

Os métodos propostos por Yetisen et al.[52] e Koh et al.[17] são mais simples, mas requerem a recalibração da aplicação antes de cada análise para poderem construir a curva de calibração adequada às condições ambientais do momento.

A abordagem aplicada na versão anterior da aplicação [37] é baseada no método proposto por Mutlu et al.[24], mas inclui algumas alterações com o propósito de o tornar viável numa aplicação móvel. Foi testado um maior número de modelos de cor e classificadores diferentes, de modo a obter o que se melhor adaptava a análise da glucose. O facto de os modelos serem treinados previamente fora da aplicação também é importante, pois permite remover esta parte da computação mais pesada da aplicação móvel, tornando o processo de análise mais rápido. Os resultados favoráveis apresentados para a análise da glucose indicam que esta abordagem pode possivelmente ser aplicada a outras substâncias.

Apesar de não ser um método de classificação, a análise com o HCA proposta por Jia et al.[14] também demonstrou resultados favoráveis pelo que seria interessante incluí-la neste processo para visualizar os conjuntos de dados.

No contexto desta dissertação há a necessidade de analisar outra substância para além da glucose: o ácido úrico. Será então necessário aplicar ao ácido úrico a mesma abordagem de seleção de modelos de cor e classificadores, feita na versão anterior [37], para poder descobrir qual a combinação que permitirá obter uma classificação mais correta.

## 3.2 Calibração da Cor

Geralmente as imagens dos sensores colorimétricos são capturadas usando um *scanner* ou uma câmara digital profissional com uma fonte de luz fixa [14]. Quando existe uma fonte de iluminação constante, uma posição da captura fixa e um dispositivo definido, é possível processar as imagens diretamente. No entanto, apesar de oferecerem bons resultados, estes métodos não são viáveis para a maioria dos utilizadores visto que requerem equipamento e condições de iluminação específicas que nem sempre estão acessíveis.

O uso do *smartphone* como dispositivo para a análise colorimétrica traz a vantagem de este ser um meio amplamente disponível, portátil e fácil de utilizar. Contudo, o seu uso origina inclui problemas relacionados com a influência da iluminação, a distância e o ângulo a que a imagem é capturada, algo que não acontece quando se opta pelos métodos tradicionais já referidos. Estes fatores exteriores vão influenciar a cor da imagem capturada, podendo levar a resultados menos corretos nas análises.

Para além das influências ambientais, o uso de diferentes modelos de *smartphones* pode ter impacto nos resultados obtidos. Fatores como o próprio sensor da câmara e os algoritmos de correção de cor aplicados automaticamente às fotografias fazem com que haja variação nas imagens obtidas [14].

Nesta secção será feita uma descrição de algumas abordagens que tentam colmatar a influências dos fatores descritos no processo da análise colorimétrica, bem como a abordagem proposta na versão anterior da aplicação.

### 3.2.1 Acessórios para o Smartphone

Uma das abordagens utilizadas em alguns dos trabalhos analisados é o uso de adereços que, quando acoplados ao *smartphone*, permitem eliminar a influência destes fatores externos ou adicionar funcionalidades que só a câmara não é capaz de fornecer.

São utilizados acessórios simples (Figura 3.4a) onde é inserida a amostra para a isolar da influência da luz e evitar diferenças na posição em relação à câmara [25, 36]. Também existem adereços mais complexos com fontes de luz integradas (Figura 3.4b) que as utilizam não só para obter condições de iluminação controladas mas também para ativar as próprias reações que pretendem analisar [3, 7, 9].



(a) Inserção do sensor no sistema para realizar análise do pH do sensor [25]



(b) Plataforma desenvolvida em [7] para testes colorimétricos

Figura 3.4: Exemplos de acessórios para o *smartphone*

Apesar de praticamente eliminar os problemas apresentados, a introdução de acessórios adicionais algo complexos, retira em parte o fator da acessibilidade e portabilidade que se procura com o uso do *smartphone* para realizar estas análises.

### 3.2.2 Métodos de Compensação

Quando não existem objetos que possibilitem a restrição ou o controlo da luz ambiente, é preciso recorrer a algoritmos que permitam de alguma forma compensar o efeito dessa iluminação na imagem, restaurando a cor correta.

A maioria dos trabalhos diretamente relacionados com biossensores de papel usam abordagens semelhantes que se baseiam no desenho de marcas no próprio sensor. Essas marcas são depois utilizadas para efetuar operações de correção de cor, escala ou rotação. Foi então necessário procurar métodos pertencentes a áreas de aplicação diferentes que, usando o *smartphone* para a captura de imagens, mantivessem o objetivo de corrigir a cor da imagem de forma eficiente.

Assim são apresentados de seguida projetos inseridos em três áreas diferentes: biossensores, realidade aumentada e câmaras digitais. São também descritos alguns projetos que, mesmo não sendo aplicados a uma área específica, apresentam abordagens diferentes como o uso da cor relativa e o modelo de cor CIE Lab.

#### 3.2.2.1 Biossensores

Na abordagem proposta por Lopez-Ruiz et al. [20] a correção da distância e do ângulo da imagem é feita com recurso a duas marcas pretas de formas distintas no sensor presente na figura 3.5a. Estas ajudam o utilizador a perceber se está a fotografar o dispositivo na posição correta. Após detetar a localização das marcas são diretamente aplicadas transformações à imagem que corrigem a rotação e a escala da mesma.

Para colmatar as diferenças de iluminação recorre-se ao uso da lanterna do *smartphone* como fonte de luz para obter mais controlo sobre a iluminação ambiente. As definições da função de *white-balance* da câmara são alteradas para que a temperatura da cor desta

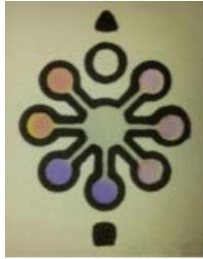
função seja semelhante à da luz emitida pela lanterna. Assim, são obtidas imagens com menor diferença de cor entre si, sendo preciso apenas aplicar uma normalização da imagem através de uma zona branca de referência do sensor para obter uma correção de cor final. Esta operação é aplicada após extrair os valores RGB de uma das zonas em análise e é dada pela fórmula 3.6, sendo  $n$  o número de bits usado pelo *smartphone* para codificar cada canal de cor.

$$RGB_{normalized} = \frac{2^n RGB_{measured}}{RGB_{white}} \quad (3.6)$$

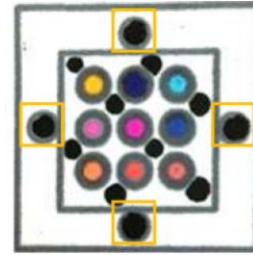
No método proposto por Jia et al. [14] é utilizado o sensor da figura 3.5b com vários poços rodeados de uma moldura branca na qual existem quatro círculos pretos. Estes círculos pretos e a zona branca em seu redor são usados como referência para os cálculos que permitem calibrar a cor da imagem. O valor de referência preto ( $R_b$ ) é a média dos valores RGB dos quatro pontos presentes no sensor, enquanto que a referência branca ( $R_w$ ) é a média dos valores RGB da zona branca próxima desses pontos pretos. Por fim é aplicada a toda a imagem inicial ( $R_{meas}$ ) a transformação descrita na equação seguinte:

$$R_{corr} = \left( \frac{256}{R_w - R_b} \right) (R_{meas} - R_b) \quad (3.7)$$

Foram ainda realizados testes recorrendo à aplicação de um algoritmo de *clustering* hierárquico HCA que obtiveram resultados favoráveis, tanto no âmbito das diferenças de iluminação entre vários *smartphones* e ambientes, mas também quando se verifica variação do ângulo ou distância de captura da imagem.



(a) Sensor com as duas marcas de localização distintas e sete áreas de amostra [20]



(b) Sensor com as quatro marcas de referência assinaladas e nove áreas de amostra [14]

Figura 3.5: Biossensores com marcas para calibração.

### 3.2.2.2 Realidade Aumentada

A Realidade Aumentada é uma tecnologia que permite a projeção de objetos gerados virtualmente sobre objetos físicos existentes em tempo real [55]. Existem vertentes desta tecnologia que usam marcadores para calcular a localização e direção da câmara, permitindo a projeção do conteúdo desejado. Estes marcadores contêm padrões simples (coloridos ou não), mas distintos de objetos comuns para que possam ser facilmente reconhecidos e processados na cena observada. Como estamos perante a necessidade de



processar padrões simples de forma eficiente maioritariamente através de dispositivos móveis, podemos estabelecer um paralelo entre este processamento e o do biossensor.

Lee e Woo [19] aborda a correção de cor para marcadores de padrões coloridos. O processo efetuado é o seguinte:

1. **Encontrar os *pixels* da região preta  $R_b$  e da região branca  $R_w$**  através da aplicação de algoritmos de componentes ligados. Os componentes ligados são regiões de *pixels* adjacentes com o mesmo valor [47].
2. **Aplicar operações de erosão** nas regiões detetadas para evitar erros no cálculo das fronteiras. Estas operações reduzem as fronteiras de um objeto, tornando o mais estreito. À medida que o filtro de erosão passa pela imagem, é calculado o mínimo local da área do filtro e o pixel onde o filtro está colocado é substituído por esse valor mínimo [47].
3. **Estimar o ponto preto  $B^e(b_R^e, b_G^e, b_B^e)$  e o branco  $W^e(w_R^e, w_G^e, w_B^e)$**  sob a iluminação desconhecida.

Assumindo que as cores dos *pixels* nas regiões  $R_b$  e  $R_w$  seguem uma distribuição normal, são criados os dois modelos  $N_b(m_b, \sigma_b)$  e  $N_w(m_w, \sigma_w)$  correspondentes, sendo  $B^e$  e  $W^e$  são definidos pela equação seguinte:

$$B^e = m_b + \Gamma \cdot \sigma_b \quad W^e = m_w + \Gamma \cdot \sigma_w \quad (3.8)$$

Onde  $\Gamma$  é um vetor que transforma um vetor de cor na iluminação atual em outro numa iluminação canónica. Assim, para encontrar  $B^e$  e  $W^e$  ótimos para a correção da cor, é preciso encontrar o  $\Gamma(\gamma_R, \gamma_G, \gamma_B)$  que minimiza o custo de mapear  $B^e$  para o valor RGB preto (0,0,0) e  $W^e$  para o valor RGB branco (255,255,255).

Para cada canal RGB, tendo em conta o ruído presente na cor que impede a correspondência total entre a cor resultante da correção  $I_f$  e a cor real esperada  $I_{f_d}$ , o custo de mapear um *pixel* é dado por:

$$C(I_f) = \|I_{f_d} - I_f\|^2 \quad (3.9)$$

Após calcular os custos médios de mapear todos os *pixels* das duas regiões para os valores desejados podemos obter o custo total do mapeamento para cada um dos canais através da soma destes. Podemos assim estimar  $\gamma_R, \gamma_G$  e  $\gamma_B$  e finalmente obter  $B^e$  e  $W^e$ .

4. **Calcular a correção de cor a aplicar usando  $B^e$  e  $W^e$**

Com esses fatores calculados e sendo  $P(I_R, I_G, I_B)$  o *pixel* da imagem original e  $P_f(I_{f_R}, I_{f_G}, I_{f_B})$  o *pixel* corrigido podemos aplicar a correção da cor através das seguintes equações:

$$I_{f_R} = 255 \left( \frac{I_R - b_R^e}{w_R^e - b_R^e} \right) \quad I_{f_G} = 255 \left( \frac{I_G - b_G^e}{w_G^e - b_G^e} \right) \quad I_{f_B} = 255 \left( \frac{I_B - b_B^e}{w_B^e - b_B^e} \right) \quad (3.10)$$

A aplicação deste processo teve bons resultados na correção da imagem do marcador sob diferentes tipos de iluminação, tal como pode ser visto pela figura 3.6. As vantagens deste algoritmo são o facto de as cores usadas em todo o processo serem provenientes da própria imagem (não necessita de referências) e também a correção ser aplicada apenas nas regiões coloridas em vez de na totalidade da imagem, o que origina um processamento mais rápido.

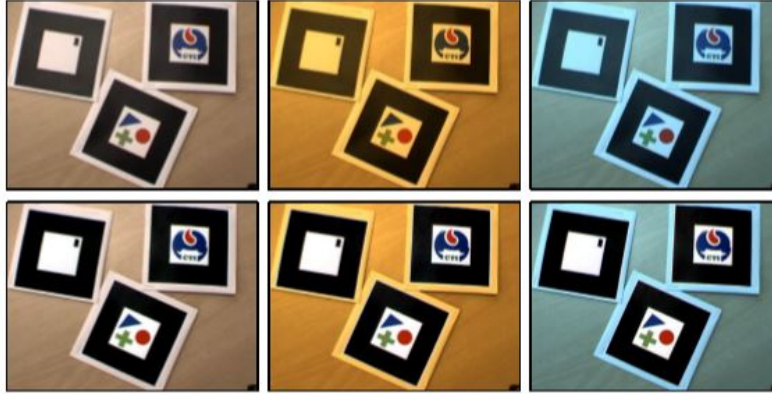


Figura 3.6: Resultados da correção de cor de marcadores sob iluminações diferentes [19]

### 3.2.2.3 Câmaras Digitais

Um dos processos mais importantes realizados numa câmara digital é o mapeamento do espaço de cores específico do seu sensor para um espaço de cor standard baseado em CIE XYZ. Este processo tem dois passos.

No primeiro passo é feita uma operação de *white-balance* para remover os efeitos da iluminação na imagem. De seguida é necessário aplicar uma transformação do espaço de cor (CST) para fazer o mapeamento dos valores resultantes da operação anterior para um espaço de cor *standard* [15]. Este processo baseia-se em dois CST já calculados na fábrica e presentes nas definições da câmara que correspondem a duas condições de iluminação fixas com cores de temperatura (CCT) muito diferentes para poderem representar o maior espectro possível de iluminação.

Quando se tira uma fotografia, a sua iluminação estimada é usada para calcular o valor da temperatura de cor da iluminação. Com base nessa temperatura, é feita uma interpolação entre os CST pré-definidos para obter o CST final que é depois aplicado na imagem da seguinte forma:

$$T_l = gT_{l_1} + (1 - g)T_{l_2} \quad \text{com} \quad g = \frac{CCT_1^{-1} - CCT_{l_2}^{-1}}{CCT_{l_1}^{-1} - CCT_{l_2}^{-1}} \quad (3.11)$$

Onde  $T_{l_1}$  e  $T_{l_2}$  se referem aos CST das duas iluminações previamente fixas e  $g$  corresponde ao peso da interpolação. Quando a imagem capturada não corresponde a nenhuma

das iluminações fixas, a interpolação irá dar origem a uma imagem com menor precisão na cor.

Em [15] são propostas duas melhorias ao método de interpolação tradicional. O primeiro método proposto é um processo de *white-balance* e consiste na extensão do método normal para incluir uma iluminação calibrada adicional no processo. A nova iluminação tem um CCT a 5000K, ficando a meio da escala da temperatura das cores. Quando se obtém uma imagem nova, é estimada a iluminação da cena e escolhido o par de CST pré-calibrados que serão interpolados baseado no CCT da iluminação estimada. O CST final é calculado através das equações 3.11. A adição desta iluminação pré-calibrada tem bons resultados e é fácil de implementar no processo existente da câmara.

O segundo método é um processo de *full color balance* e recorre a algoritmos de aprendizagem automática para obter a correção de todas as cores. Neste caso não é necessário haver um CST para cada iluminação, podendo-se apenas calcular um CST fixo. A estimativa deste CST fixo pode ser feita através de uma única imagem do quadro de referência para uma certa iluminação (6500K) ou de uma imagem do quadro para cada iluminação. Sendo que para cada imagem usada é preciso estimar uma matriz de correção de cores total através de um classificador de *Bayes* [6].

Os resultados obtidos para câmaras de telemóveis destacados na figura 3.7 revelaram que a abordagem do CST fixo com várias observações apresenta os melhores resultados, seguida da mesma abordagem para uma única observação e por fim a descrita no primeiro método proposto. As três soluções propostas obtêm melhores resultados que os métodos básicos da câmara. Apesar de obter melhores resultados, o segundo método recorre a algoritmos de *aprendizagem automática*, o que o torna mais pesado em termos de computação do que o primeiro método. Assim, pode não ser o mais indicado para dispositivos móveis.

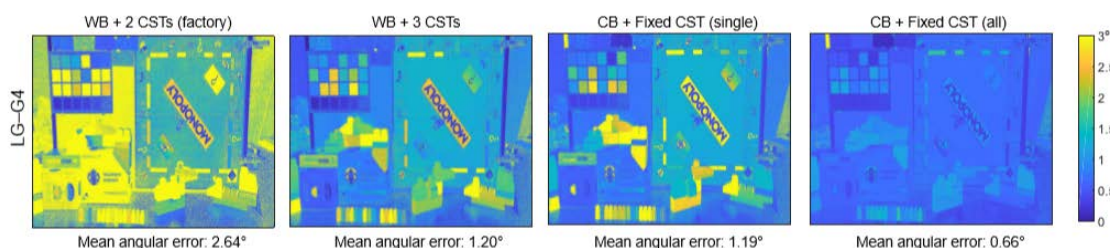


Figura 3.7: Representação visual do erro da reprodução da cor em imagens capturadas com um *smartphone* LG-G4 [15]

#### 3.2.2.4 Outros Métodos de Compensação

**Modelo CIE Lab** Em [54] foi desenvolvida uma técnica de microscopia com a capacidade de obter imagens com alta fidelidade de cor e resolução através de um sistema de microscópios portátil e de baixo custo.

O microscópio sem lente capta várias imagens de uma mesma zona com pequenas alterações entre si e utiliza-as para reconstruir uma imagem em escala de cinzento de alta

resolução, enquanto que o microscópio do *smartphone* captura uma colorida, mas de menor resolução. Estas duas imagens são depois combinadas para se obter uma imagem de cor precisa com elevada resolução. Dada a natureza das imagens capturadas, a calibração tem um papel fundamental nesta abordagem.

O objetivo desta calibração é encontrar um mapeamento que corrija as cores capturadas por um microscópio digital a partir das cores definidas numa referência. Esta é construída a partir de amostras de um quadro de referência normal, mas tem uma escala muito menor (0,5 mm a 1 mm), pelo que se recorre a um microscópio sem lente para obter uma imagem hiperespectral da mesma [54].

O processo desenvolvido tem quatro passos e só precisa ser feito uma vez por sistema, isto é, uma vez por cada combinação dos aparelhos usados. Assim as funções resultantes de cada passo são guardadas no sistema para que possam ser aplicadas noutra situação.

1. **Normalização da imagem para *white balance*** A imagem da amostra é dividida por uma imagem de calibração sem amostra. Após a normalização a imagem é convertida de RGB para o espaço de cor CIE Lab, descrito na secção 2.1.1, onde L representa a luminosidade e (a,b) a cor.
2. **Correção da luminosidade** Para corrigir a luminosidade calcula-se uma função de regressão entre as componentes L do resultado da normalização e da referência.
3. **Correção da ausência de saturação** Neste passo, a saturação da imagem capturada é alterada para corresponder à saturação das cores de referência, através da amplificação do seu componente da cor ( $c = \sqrt{a^2 + b^2}$ ). Para isso, calcula-se uma função entre a cor do resultado obtido no passo anterior e a cor da referência que é depois utilizada para alterar as componentes a e b do resultado obtido no passo dois.
4. **Transformação da cor** Para obter a matriz de transformação de cor final recorre-se a uma regressão polinomial [54] a partir dos componentes do resultado obtido no passo anterior e os componentes de referência. No final do processo a imagem volta a ser transformada para RGB.

Para testar o processo de calibração, foram calibrados dois microscópios óticos, um de *smartphone* e outro normal. Como se pode observar na figura 3.8, antes da calibração as cores capturadas eram muito diferentes dos valores de referência e havia bastante ausência de saturação, sendo que depois da calibração praticamente todas as suas cores corresponderam com as de referência. As cores das imagens obtidas foram precisas e bastante próximas das capturadas com um microscópio comum, o que revela o potencial deste método.

**Cor Relativa** Em [44] é proposta uma nova abordagem para análise espectroscópica realizada a partir do *smartphone* e sem calibração prévia, através da determinação da cor relativa em vez da absoluta. Foi criada uma aplicação para analisar a intensidade relativa

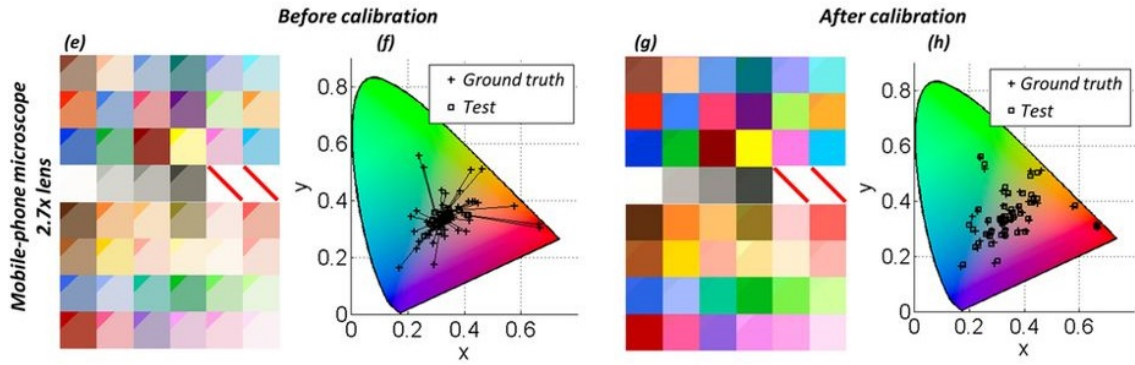


Figura 3.8: Resultados da calibração da cor

(e) Cores capturadas pelo microscópio do *smartphone* com ampliação de 2,7x. O triângulo no canto superior esquerdo de cada quadrado é a cor de referência extraída e o resto é a cor capturada pelo microscópio.

(f) Cores em (e) descritas no diagrama do espaço de cor CIE XYZ, onde + representa a cor de referência e □ representa a cor obtida.

(g,h) cores após a calibração [54].

da cor de duas regiões diferentes da mesma imagem, sendo também explorada a análise de absorção de moléculas em nanopartículas metálicas.

Mesmo com iluminação constante, a mesma imagem obtida com diferentes dispositivos pode ter cores RGB distintas. Assim, para fazer uma análise com qualidade é preciso realizar uma calibração com recurso a um objeto de referência cujos valores RGB sejam já conhecidos. A cor em RGB dada pelo dispositivo pode ser descrita como:

$$R' = R^\gamma, G' = G^\gamma, B' = B^\gamma$$

. Onde  $R$  é componente da cor real,  $R'$  da cor da imagem obtida e  $\gamma$  um fator de correção. Se vários dispositivos tiverem aproximadamente o mesmo  $\gamma$ , a sua percepção da cor será semelhante, mas se forem diferentes, o valor das cores obtidas será distinto.

Deste modo, os parâmetros de cor relativos ( $r, g, b$ ) são definidos como:

$$r = \frac{\ln(R'_i)}{\ln(R'_r)} = \frac{\ln(R_i)}{\ln(R_r)}, \quad g = \frac{\ln(G'_i)}{\ln(G'_r)} = \frac{\ln(G_i)}{\ln(G_r)}, \quad b = \frac{\ln(B'_i)}{\ln(B'_r)} = \frac{\ln(B_i)}{\ln(B_r)} \quad (3.12)$$

Onde  $i$  é a região de interesse e  $r$  a de referência. As cores relativas não dependem de  $\gamma$ , só das cores reais do objeto.

Para além do método proposto acima foram explorados mais três outros métodos para obter a cor relativa propostos por outros autores para estabelecer uma comparação [44]. Para testar o método proposto foram criados dois padrões: um com quatro áreas de cores diferentes (vermelho, verde, azul e branco) e outro com 15 zonas de vários tons de vermelho. As cores verdadeiras dos padrões foram determinadas usando um colorímetro, sendo depois fotografados com quatro *smartphones* diferentes.

O primeiro padrão foi usado para testar as diferenças de cor ao usar *smartphones* diferentes, sendo a cor branca usada como referência face aos outros componentes. O

método proposto apresentou o desvio padrão mais baixo para todas as componentes RGB (2%, 3% e 3.5%), o que indica que este depende pouco do dispositivo e reduz as discrepâncias entre as cores obtidas em dispositivos diferentes.

O segundo padrão foi usado para testar as diferenças de cor provocadas por quatro tipos de iluminação diferentes (fluorescente, incandescente, LED branco e solar), sendo que o método proposto voltou a obter o desvio padrão mais reduzido, o que indica uma menor dependência da iluminação. Analisou-se ainda a capacidade de distinção entre diferentes tons da mesma cor. Constatou-se que este método consegue distinguir 2% de diferença entre tons na cor vermelha.

### 3.2.3 Versão Anterior da Aplicação

Na versão anterior do projeto [37] foi aplicada uma abordagem baseada no uso de cores de referência. No entanto, em vez de usar um quadro de referência em conjunto com o sensor, as quatro cores de referência (branco, azul, verde e vermelho) foram inseridas no próprio sensor, rodeando o poço da amostra.

O método consiste na aplicação de uma matriz de correção de cor  $M$  a todos os *pixels* da imagem a partir de um sistema de equações da forma  $A \times M = B$  em que se procura minimizar  $|A \times M| - |B|$ .

$$A = \begin{bmatrix} R_b & G_b & B_b \\ R_g & G_g & B_g \\ R_r & G_r & B_r \\ R_w & G_w & B_w \end{bmatrix} B = \begin{bmatrix} R'_b & G'_b & B'_b \\ R'_g & G'_g & B'_g \\ R'_r & G'_r & B'_r \\ R'_w & G'_w & B'_w \end{bmatrix} M = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \quad (3.13)$$

Sendo que  $(R_i, G_i, B_i)$  são as cores da imagem extraídas com o primeiro momento da cor e  $(R'_j, G'_j, B'_j)$  as cores de referência, com  $i \in \{r, g, b, w, p\}$  e com  $j \in \{r, g, b, w\}$ . Neste caso,  $r$  corresponde à zona de calibração vermelha,  $g$  à verde,  $b$  à azul e  $w$  à branca, sendo  $p$  a cor do poço. Dada a cor do poço  $(R_p, G_p, B_p)$ , a cor calibrada  $(R_p^f, G_p^f, B_p^f)$  será dada por:

$$[R_p^f, G_p^f, B_p^f] = [R_p, G_p, B_p] \times M \quad (3.14)$$

As cores de referência utilizadas foram escolhidas com base nos momentos de cor de cada zona de referência extraídos de imagens de sensores digitalizadas, aproximando assim a cor capturada da cor das imagens usadas no passo da análise colorimétrica.

De acordo com testes desenvolvidos, a calibração utilizada não foi suficiente face a imagens capturadas em diferentes câmaras e em baixa iluminação. Estas tiveram uma influência negativa no processo da análise colorimétrica do sensor.

### 3.2.4 Análise Crítica

Os trabalhos que envolvem o uso de adereços descritos na secção 3.2.1 eliminam praticamente os problemas com a influência da iluminação e de outros fatores externos. Infelizmente, a introdução de acessórios no *smartphone* retira em parte a acessibilidade e

a portabilidade que se procura, visto que estes podem ser complexos e não adaptáveis a vários modelos de *smartphone*.

A proposta de Lopez-Ruiz et al.[20] de usar a lanterna como fonte de luz para tentar obter o controlo deste fator é simples tanto em termos de processamento como de alterações no sensor. No entanto, não foram testados muitos *smartphones* diferentes e, conseqüentemente, não são conhecidos os efeitos do uso da lanterna e das funções de correção automática da câmara nos resultados usando *smartphones* com câmaras de gamas distintas.

O método proposto por Jia et al.[14] é um processo de processamento simples, mas que requer uma disposição característica do sensor, pelo que seria necessário alterar o design atual. No entanto, apresenta resultados favoráveis de calibração do sensor, testados em modelos de *smartphones* diferentes. Tal como o anterior, o método de Lee e Woo[19] também depende apenas das áreas brancas e pretas do marcador para efetuar os cálculos de correção, se bem que estes são mais complexos. Como as alterações no sensor necessárias para reproduzir estes métodos não são muito complexas, podemos considerá-los abordagens a avaliar.

As propostas de Zhang et al.[54] e Souza et al.[44] são também interessantes, pois apresentam métodos baseados em técnicas diferentes das anteriores. Zhang et al. propõe um processo de correção da luminosidade e da saturação através de cálculos efetuados sobre o modelo CIE Lab, enquanto que Souza et al. propõe um método para obter a cor relativa que apresenta resultados favoráveis ao lidar com dispositivos, tons e iluminações diferentes.

Por outro lado, os processos descritos em Karaimer e Brown[15] são mais complexos que os já referidos. Apesar de obterem bons resultados para as correções nas câmaras de *smartphone*, requerem o uso de um quadro de referência e acesso a imagens intermédias entre processos automáticos da câmara do *smartphone*, algo a que não é possível aceder através da aplicação.

A calibração da versão anterior baseava-se em cálculos feitos a partir de quatro zonas de referência de diferentes cores desenhadas no próprio sensor. De acordo com os testes desenvolvidos, este processo não foi o suficiente para que a aplicação conseguisse lidar com as imagens capturadas em diferentes câmaras e em baixa iluminação, que tiveram uma influência negativa no resultado da análise.

Por esta razão é necessário testar as novas abordagens apresentadas, de modo a perceber quais podem ser potencialmente aplicadas em versões futuras de biossensores deste género.





## MÉTODOS

Neste capítulo são descritos os processos para o estudo e seleção de algoritmos para a análise colorimétrica e para a calibração da cor.

No que diz respeito à análise colorimétrica é feita uma descrição do processamento dos dados e do treino e validação dos vários classificadores estudados. Em relação à calibração, inicialmente é apresentada uma modificação proposta para o algoritmo usado na versão anterior da aplicação, sendo de seguida descrito o processo de teste e comparação entre os vários algoritmos com o objetivo de selecionar aquele que proporcione melhores resultados em termos da consistência da cor.

### 4.1 Análise Colorimétrica

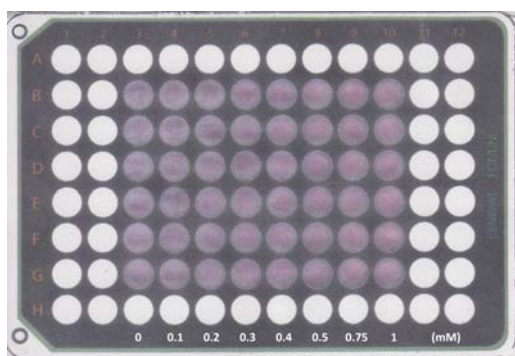
No estudo da análise colorimétrica foi reproduzido o processo realizado no trabalho anterior [37], onde foram testados vários classificadores, combinações de *features* e modelos de dados diferentes para escolher a melhor solução para a classificação da glucose.

Com a inclusão de uma nova substância na aplicação, o ácido úrico, foi necessário reproduzir o processo utilizado na análise da glucose para escolher um classificador que se adapte a este novo problema. Em relação à glucose, pretende-se averiguar se a utilização de um maior conjunto de dados contribui para melhores resultados.

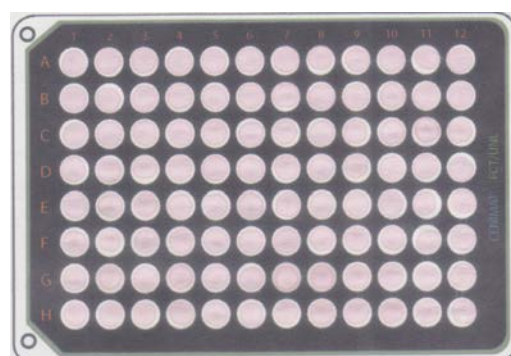
O processo inicia-se com a extração e processamento dos dados, que inclui uma análise da cor para identificar combinações de *features* apropriadas para os dois problemas. De seguida é feito o treino e validação dos vários classificadores para obter resultados que permitam fazer uma comparação e decidir qual o melhor algoritmo para implementar na aplicação. Todo o processamento dos dados, escolha de parâmetros e treino dos classificadores são feitos com recurso a funções da biblioteca *Scikit-Learn* para a linguagem de programação *Python* [32].

### 4.1.1 Obtenção e Processamento de Dados

O treino de classificadores necessita de dados representativos daquilo que pretendemos classificar. Foram fornecidas pelo CENIMAT diversas imagens de placas referentes ao ácido úrico (figura 4.1a) e à glucose (figura 4.1b). Estas placas contêm várias zonas de cor que representam a reação química do biossensor para uma certa concentração da substância. Através destas imagens é possível extrair a informação necessária para fazer a classificação destas substâncias.



(a) Placa de dados de ácido úrico. Cada coluna corresponde a uma concentração em mM (0;0,1;0,2;0,3;0,4;0,5;0,75;1).



(b) Placa de dados de glucose correspondente à concentração de 7 mM.

#### 4.1.1.1 Conjuntos de Dados

No trabalho anterior [37] foi proposta a melhoria dos resultados através da expansão do conjunto de dados usado para análise da glucose. Assim, foram fornecidas pelo CENIMAT novas repetições para algumas das concentrações com vista a equilibrar o número de exemplos por cada classe em alguns modelos.

No entanto, como se pode observar na figura 4.2, foram detetados problemas com a consistência da cor entre as repetições de algumas concentrações, nomeadamente as de 15 mM e 20 mM. Optou-se então por não introduzir estas novas repetições no novo conjunto de dados, pois estas iriam prejudicar os resultados da classificação, sendo adicionadas então as novas amostras de 1,5 mM, 5 mM, 7 mM e 10mM aos dados usados anteriormente. Segundo o CENIMAT, existem três situações que podem ter influenciado negativamente a consistência da cor. A variação dos volumes e concentrações dos reagentes usados na reação, a variação das condições ambientais (temperatura e humidade) no momento dos ensaios e ainda variação nos tempos do ensaio. Nomeadamente os tempos de colocação do reagente no papel, os tempos de secagem e os tempos entre a reação e a captação da imagem através do *scanner*. Estas podem ser combatidas através de um ambiente e procedimento mais controlado [33].

O número de observações existentes para cada concentração está descrito na tabela 4.1, para a glucose e na tabela 4.2 para o ácido úrico .

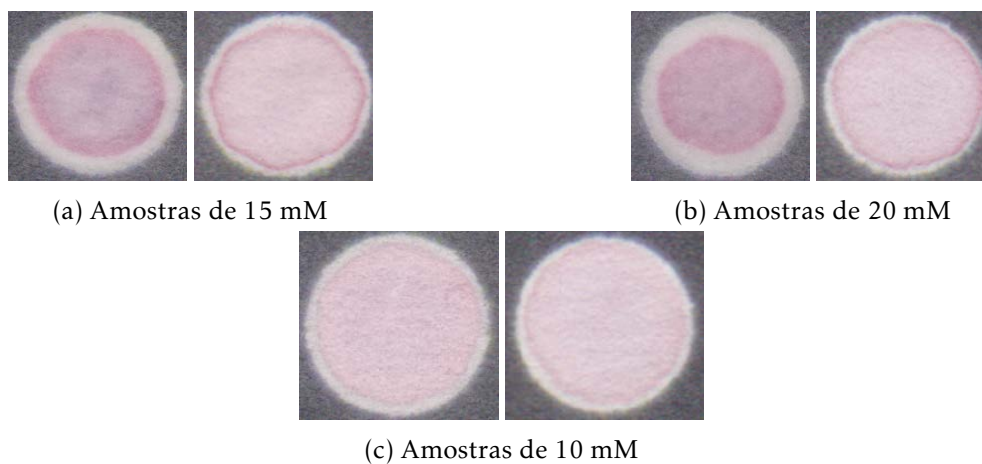


Figura 4.2: Comparação entre as amostras do conjunto de dados inicial usado no trabalho anterior (esquerda) e as novas amostras fornecidas para esta dissertação (direita). As amostras de 15 mM (a) e 20 mM (b) não são consistentes. As restantes concentrações como, por exemplo, os 10 mM (c), são consistentes.

Concentração (mM)	Nº de amostras atuais	Nº de amostras na versão anterior
0	56	24
1.25	23	23
1.5	120	24
2	24	24
5	120	23
7	120	24
10	120	24
12.5	24	24
15	23	23
20	24	24

Tabela 4.1: Número de amostras por cada concentração no conjunto de dados usado para a glucose em comparação com os dados usados na versão anterior da aplicação [37].

Concentração (mM)	Nº de observações
0	48
0.1	48
0.2	48
0.3	48
0.4	48
0.5	48
0.75	48
1	48

Tabela 4.2: Número de observações por cada concentração no conjunto de dados usado para construir os modelos do ácido úrico.

Para além do estudo das concentrações, foram também analisadas diferentes disposições dos dados por níveis de referência.

No trabalho anterior [37], devido aos resultados insuficientes obtidos na classificação das concentrações da glucose, foi criada uma disposição dos dados de acordo com quatro níveis de referência desta substância. Esta disposição divide-se ainda em duas situações que influenciam o significado da concentração da glucose para o utilizador: jejum e duas horas após a refeição, dependendo do seu estado quando realiza a análise.

Apesar de fornecer melhores resultados que a concentração, este modelo não maximiza a informação que poderia providenciar ao utilizador. Tal deve-se ao facto de uma das classes juntar dois dos níveis de referência num só, tornando o resultado pouco claro. Como estamos perante um maior conjunto de dados para esta substância do que o usado anteriormente, optou-se por analisar também os dados com os cinco níveis de referência, de modo a procurar obter a informação mais útil possível para o utilizador. As cardinalidades destes dois modelos encontram-se nas tabelas 4.3 e 4.4, respetivamente.

Nível	Nº de observações	
	Em jejum	Duas horas após a refeição
Vazio	56	56
Hipoglicemia	167	167
Normal/Pré-Diabetes	240	360
Diabetes	191	71

Tabela 4.3: Número de observações por cada classe do modelo de dados dos quatro níveis de referência da glucose, em jejum e duas horas após a refeição.

Nível	Nº de observações	
	Em jejum	Duas horas após a refeição
Vazio	56	56
Hipoglicemia	167	167
Normal	120	240
Pré-Diabetes	120	120
Diabetes	191	71

Tabela 4.4: Número de observações por cada classe do modelo de dados dos cinco níveis de referência da glucose, em jejum e duas horas após a refeição.

Em relação ao ácido úrico, segundo o verificado anteriormente na glucose, para além da análise das concentrações fornecidas, optou-se por testar logo à partida um modelo dividido pelos três níveis.

Ao contrário da glucose, existem níveis de referência diferentes para os sexos masculino e feminino [33]. No entanto, devido à reduzida diferença entre os intervalos de concentrações dos dois sexos e como se trata de uma primeira análise, optou-se por incluir apenas um dos modelos no estudo, nomeadamente o do sexo masculino. O número de

observações dos vários níveis está descrita na tabela 4.5.

Nível	Nº de observações
Baixo	96
Normal	144
Elevado	144

Tabela 4.5: Número de observações verificadas para cada classe do modelo de dados dos três níveis de referência do ácido úrico.

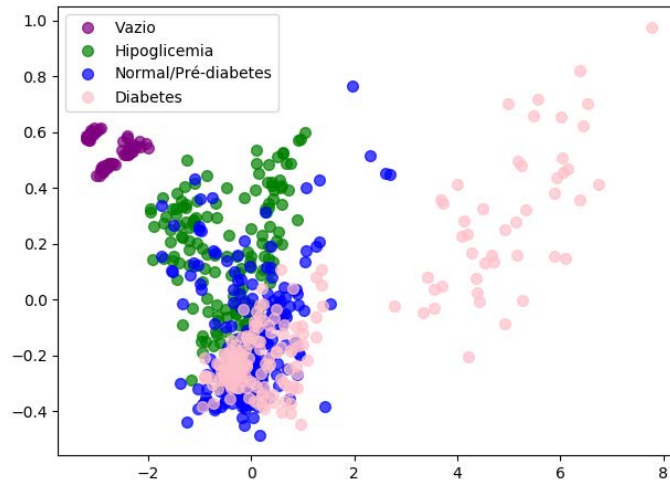
#### 4.1.1.2 Extração dos Dados

A partir das imagens das placas é feita a extração da cor que corresponderá a uma certa concentração, através de um programa desenvolvido em *Python* com funcionalidades da biblioteca *OpenCV*. Este programa segue um processo semelhante ao aplicado na deteção do biossensor pela aplicação, descrito na secção 6.1.

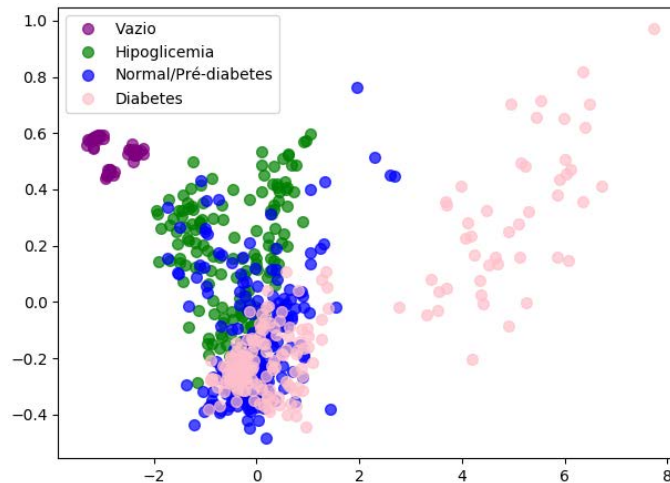
Os círculos da placa são detetados através da transformada *Hough Circles* e a informação é extraída de uma região interior do círculo correspondente a 60% do seu raio. Seleccionamos apenas esta parte do círculo para evitar a inclusão da zona preta circundante, que pode influenciar o valor da cor.

No trabalho anterior [37] foi extraída apenas a média da cor da região, valor este que foi convertido para vários espaços de cor diferentes e usado para treinar os modelos. Sendo um dos objetivos melhorar os resultados da classificação, procurou-se adicionar outras características, para além da média, que pudessem fornecer mais informação ao processo. Stricker e Orengo [45] recorrem aos três primeiros momentos da cor: média, desvio padrão e assimetria para descrever a imagem, enquanto Musto et al. [23] usam a diferença de cor entre o sensor antes e após da exposição a um reagente. Assim, para além da média, foram extraídos ainda o desvio padrão e a assimetria da região, bem como o seu histograma de cor, usado em seguida para identificar os descritores mais úteis para a classificação das substâncias. O valor da diferença entre a média da cor do sensor vazio e com alguma concentração foi extraído inicialmente. No entanto, após observação da representação gráfica dos dados (figura 4.3) concluiu-se que este valor era redundante face à observação apenas do valor da média, pelo que este descritor em particular não apresentou utilidade para estes problemas a analisar.

Os quatro descritores seleccionados são depois convertidos para os vários modelos de cor a analisar: RGB, HSV, HLS, LAB, XYZ e YUV. Por último, todos os valores são guardados num ficheiro .csv para facilitar a sua leitura tanto pelos programas das fases seguintes da análise, como por programas de folhas de cálculo como o Microsoft Excel.



(a) Visualização do PCA com a média da cor da amostra(em RGB)



(b) Visualização do PCA com a diferença entre a média da cor do sensor vazio e a média da cor com amostra (em RGB)

Figura 4.3: Comparação entre a visualização dos dados através da média da cor da amostra e a diferença entre a média da cor do sensor vazio e a média da cor com amostra com recurso a PCA. A representação através da diferença é redundante quando comparada com a média da cor.

#### 4.1.1.3 Seleção de Descritores de Cor

De modo a evitar o uso de descritores redundantes ou com pouca utilidade para a diferenciação das cores entre concentrações, foi feita uma seleção prévia dos descritores a usar. Independentemente da substância a classificar, optou-se por manter a análise dos modelos de cor aplicados à glucose no trabalho anterior: RGB, XYZ, YUV, HLS, adicionando ao processo o modelo Lab.

O autor do novo biossensor [33] sugere como descritores para o ácido úrico o rácio entre os componentes vermelho e azul do RGB e a componente H do modelo HSV, enquanto que o autor do biossensor da glucose [8] sugere para essa substância, o rácio entre os componentes vermelho e verde do modelo RGB.

Outras combinações de valores selecionadas baseiam-se na observação dos histogramas das cores, como o sugerido por Kim et al. [16]. Ao observar os histogramas entre as diferentes concentrações da mesma substância podemos verificar quais componentes apresentam maior potencial para separar as concentrações. No caso da figura 4.4, por exemplo, é possível ver uma separação mais acentuada entre as várias concentrações do ácido úrico nas componentes H e S do que na componente V, do sistema de cor HSV. Através da observação destes gráficos para os vários modelos podemos isolar componentes mais relevantes para a nossa análise. Os histogramas referentes aos vários modelos e substâncias encontram-se disponíveis no apêndice E.

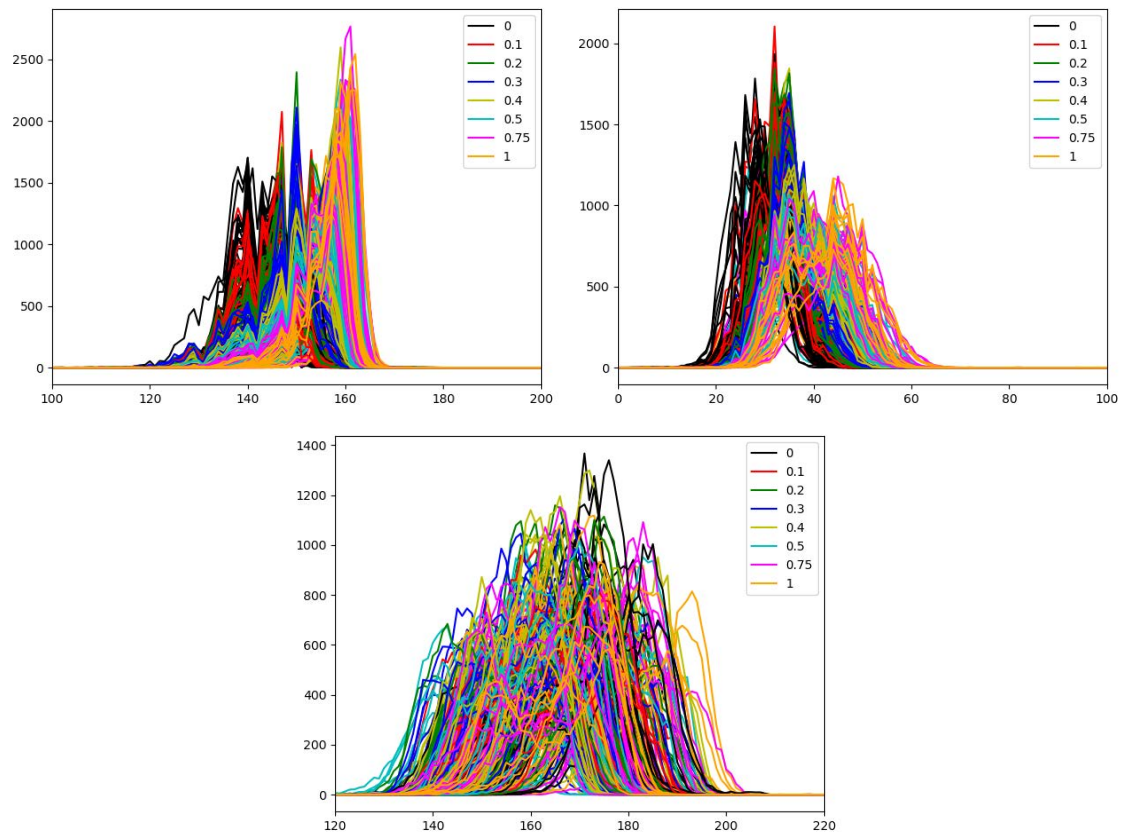


Figura 4.4: Histogramas para os três componentes HSV do conjunto de dados do ácido úrico. Cada cor do gráfico corresponde a uma das concentrações disponíveis.

Procurando ainda a separação entre as concentrações, para além da análise visual dos histogramas foram ainda observadas em maior detalhe as diferenças entre as concentrações através do *Microsoft Excel*. Neste processo, foram calculadas as somas das diferenças entre as médias dos valores das várias características para cada concentração, procurando assim obter as *features* para quais esta diferença entre as concentrações é maior.

Posto isto, as combinações de características extraídas e aplicadas nos testes com os algoritmos de classificação são as seguintes:

- Glucose

- Média em cada um dos modelos RGB, HSV, HLS, XYZ, LAB, YUV.
  - Média, desvio padrão e assimetria em cada um dos modelos RGB, HSV, HLS, XYZ, LAB e YUV.
  - Rácio R/G (RGB).
  - Média V e S (HSV).
  - Média L e S (HLS).
- **Ácido úrico**
    - Média em cada um dos modelos RGB, HSV, HLS, XYZ, LAB, YUV.
    - Rácio R/B (RGB).
    - Média H (HSV).
    - Rácio R/B (RGB) e Média H (HSV).
    - Média a e b (LAB).
    - Média H e S (HSV).
    - Média H e S (HLS).
    - Média H e S (HSV), H e S(HLS).
    - Média RGB, V(HSV), Z(XYZ), a(LAB), Y(YUV).

As representações gráficas através de PCA das diferentes disposições dos dados referidas acima, estão disponíveis no apêndice G.

#### 4.1.1.4 Processamento de Dados

Antes de iniciar o treino dos classificadores é necessário fazer o processamento dos dados extraídos, recorrendo a funções da biblioteca *Scikit-Learn* [32]. Os métodos e funções aplicadas são detalhadas na secção 2.1.3.1.

O conjunto de dados é dividido em dois conjuntos, usando a função *train\_test\_split*, que divide aleatoriamente os dados em duas partes: o conjunto de treino e o conjunto de teste. Seguindo o processo do trabalho anterior, reservamos para a glucose 67% dos dados para o conjunto de treino e 33% para o de teste. Em relação ao ácido úrico, reservou-se uma menor percentagem para o teste devido ao tamanho mais reduzido do conjunto, nomeadamente de 80% para o conjunto de treino e 20% para o de teste.

Como existem diferenças na escala dos valores da cor é preciso uniformizar os dados. Em relação à glucose foi seguido o processo estabelecido no trabalho anterior, pelo que foi aplicada a normalização comum, usando a função *StandardScaler*. No conjunto do ácido úrico foram detetados alguns valores extremos, pelo que a uniformização foi feita de acordo com a amplitude interquartil, através da função *RobustScaler*. Este método é mais resistente aos efeitos de *outliers* nos dados do que a normalização mais comum.



O passo final é a aplicação da PCA que permite reduzir a dimensionalidade dos dados com múltiplas *features* com perda de informação mínima. Este processo facilita a visualização e interpretação gráfica dos vários conjuntos. Os dados foram também analisados sem recurso ao PCA para se fazer uma comparação entre estas duas abordagens nos resultados da classificação.

#### 4.1.2 Classificação

No que diz respeito aos algoritmos de classificação testados, foi mantido o processo desenvolvido no trabalho anterior [37]. Deste modo, foram testados três tipos de classificadores: KNN, SVM e *Random Forest*, que se encontram descritos na secção 2.1.3.2.

Para assegurar que os modelos treinados estão otimizados para tirar melhor partido dos dados, antes de proceder ao treino dos algoritmos foi feita uma pesquisa exaustiva dos seus hiperparâmetros.

Caso o modelo seja muito otimizado, pode vir a adaptar-se demasiado ao conjunto de dados, perdendo a capacidade de prever a classe de amostras que não tenha observado durante o processo de treino. Esta situação é descrita como *overfitting* [1]. Para minimizar o *overfitting*, os classificadores treinados foram validados com recurso a *K-Fold-Cross Validation* de  $K = 10$  e a ainda através de previsões utilizando um conjunto de teste distinto. Estes métodos de validação estão descritos na secção 2.1.3.1.

##### 4.1.2.1 Seleção de Hiperparâmetros

Cada classificador possui parâmetros que não consegue aprender durante o treino e que por isso são definidos na sua criação (hiperparâmetros) [1]. Os hiperparâmetros de cada algoritmo usado estão descritos na secção 2.1.3.2. Assim, é preciso procurar os parâmetros que forneçam melhores resultados para o nosso problema.

Para encontrar os melhores parâmetros para cada um dos modelos e algoritmos foi aplicado o método de *GridSearchCV*. Este percorre um conjunto de parâmetros, utilizando o conjunto de treino para treinar várias instâncias do classificador a partir das quais vai procurando e validando a combinação que fornece o melhor valor para a métrica escolhida [32]. Neste contexto, em concordância com o realizado no projeto anterior, foi escolhido o *F1-Score* como métrica a melhorar. Esta métrica é baseada na precisão e na sensibilidade, dando igual importância a cada uma destas métricas no cálculo do resultado. Para evitar a escolha de parâmetros em *overfitting*, a validação do processo recorre a *K-Fold Cross-Validation*, sendo usado o valor comum de dez *folds* [1].

O processo de pesquisa é bastante intensivo, pelo que o intervalo de parâmetros a explorar foi definido de modo a ser o mais abrangente possível com o menor número de iterações. No caso do KNN é estimado apenas um parâmetro: o número de vizinhos a considerar ( $K$ ) ao escolher a classe da nova amostra [5]. Neste caso é feita uma única pesquisa no intervalo que contém todos os números ímpares de três ao número de elementos do conjunto de treino:  $\{3, 5, \dots, n\_exemplos\_treino\}$ . Como o conjunto de dados é

relativamente pequeno podemos limitar a pesquisa pelo seu total de elementos, mas caso seja muito grande, é preferível usar uma percentagem do número de elementos [11]. O uso de elementos ímpares permite evitar a existência de empates na decisão da atribuição das classes.

Os restantes classificadores utilizados requerem que sejam estimados múltiplos parâmetros. Para evitar que esse processo seja muito lento, a pesquisa foi dividida em duas partes. A primeira pesquisa é mais abrangente e percorre um intervalo grande de parâmetros, permitindo identificar uma região que forneça melhores resultados. A segunda cobre um intervalo de valores mais detalhado dentro dessa mesma região [21].

Em relação ao classificador SVM, procuramos otimizar os parâmetros  $C$ , um coeficiente de regularização, e  $\gamma$ , que influencia a capacidade do modelo de se adaptar ao formato do conjunto de dados [1, 32]. Os intervalos de parâmetros usados nas pesquisas estão detalhados na tabela 4.6, onde  $C_a$  e  $\gamma_a$  correspondem aos valores destes parâmetros que obtiveram o melhor resultado na pesquisa abrangente.

Hiperparâmetro	Abrangente	Restrita
$C$	$\{2^0, 2^1, \dots, 2^{10}\}$	$\{2^{(C_a-1)}, 2^{(C_a-1)+0.5}, \dots, 2^{(C_a+1)}\}$
$\gamma$	$\{2^{-6}, 2^{-5}, \dots, 2^6\}$	$\{2^{(\gamma_a-1)}, 2^{(\gamma_a-1)+0.5}, \dots, 2^{(\gamma_a+1)}\}$

Tabela 4.6: Lista de parâmetros usados para as pesquisas abrangente e restrita com o algoritmo SVM.

Por último, para o algoritmo *Random Forest* existem múltiplos parâmetros que podem ser otimizados. Neste contexto, procuramos otimizar: o número de árvores usadas ( $n\_estimators$ ), a profundidade máxima dessas árvores ( $max\_depth$ ), o número de *features* a considerar ao procurar a melhor divisão ( $max\_features$ ) e o número mínimo de exemplos a necessário para dividir um nó da árvore ( $min\_samples\_split$ ) [32]. Os intervalos de parâmetros usados nas pesquisas estão detalhados na tabela 4.7, onde  $n\_est_a$ ,  $depth_a$  e  $split_a$  correspondem aos valores destes parâmetros que obtiveram o melhor resultado na pesquisa abrangente.

Com os melhores parâmetros possíveis selecionados para cada modelo e classificador podemos então treinar os classificadores e verificar quais obtêm os melhores resultados para os problemas a analisar.

#### 4.1.2.2 Treino dos Classificadores

Com os parâmetros selecionados previamente os vários classificadores são treinados e validados usando o conjunto de treino, através da aplicação de *K-Fold Cross-Validation* com dez *folds*. Em simultâneo é feita também a validação utilizando o conjunto de teste, para avaliar a capacidade dos classificadores de extrapolar resultados para dados novos não observados durante o processo de treino.

Hiperparâmetro	Abrangente	Restrita
Número de árvores ( $n\_est$ )	$\{2^0, 2^1, \dots, 2^7\}$	$\{2^{(n\_est_a-1)}, 2^{(n\_est_a-1)+0.25}, \dots, 2^{(n\_est_a+1)}\}$
Profundidade máxima das árvores ( $depth$ )	$\{2^{-6}, 2^{-5}, \dots, 2^6\}$	$\{2^{(depth_a-1)}, 2^{(depth_a-1)+0.25}, \dots, 2^{(depth_a+1)}\}$
Fração de features a considerar	$\{0.1, 0.325, 0.55, 0.775, 1\}$	$\{0.1, 0.325, 0.55, 0.775, 1\}$
Fração de amostras a considerar por nó ( $split$ )	$\{0.1, 0.2, \dots, 0.5\}$	$\{2^{(split_a-0.1)}, 2^{(split_a-0.1)+0.05}, \dots, 2^{(split_a+0.1)}\}$

Tabela 4.7: Lista de parâmetros usados para as pesquisas abrangente e restrita com o algoritmo *Random Forest*.

Durante esta avaliação foram recolhidos resultados referentes às três métricas em análise: *F1-Score*, *Recall* e *Precision*, para cada modelo. Os valores da classificação do ácido úrico são detalhados no apêndice C, enquanto que os referentes à glucose podem ser encontrados no apêndice D.

Para cada modelo treinado foram ainda desenhadas as regiões de decisão, que permitem visualizar a distribuição das várias classes após a classificação. Na figura 4.5, por exemplo, podemos observar a classificação das dez concentrações da glucose usando o algoritmo KNN. As regiões referentes aos restantes modelos estão disponíveis no apêndice F.

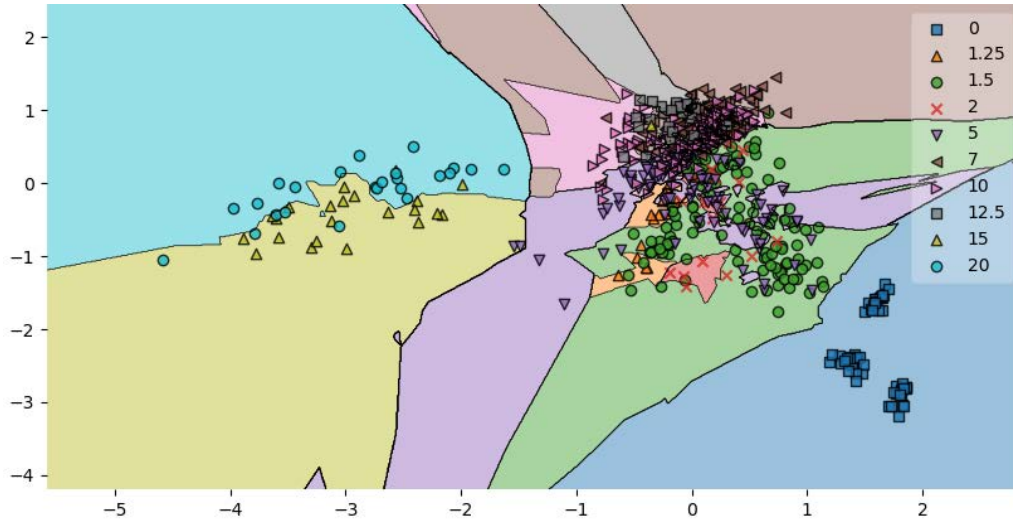


Figura 4.5: Região de decisão do modelo KNN para as dez concentrações em análise, baseado nas componentes L e S do espaço de cor HLS.

### 4.1.2.3 Seleção dos Classificadores

Sendo o objetivo obter a melhor classificação possível para cada um dos problemas é necessário decidir qual dos classificadores treinados é a melhor opção para os problemas em análise. Esta decisão deve ter em conta não só o valor da métrica principal, nomeadamente o maior *F1-Score* obtido, mas também a possibilidade de o classificador estar em situação de *overfitting*. Assim, consideramos também as diferenças entre os resultados obtidos no conjunto de validação e no conjunto de teste.

Deste modo, foi criado para a comparação um fator de seleção  $f$ :

$$f = X - Y \quad (4.1)$$

Onde  $Y$  corresponde à soma das diferenças absolutas entre os valores obtidos para cada métrica (*F1-Score*, *Precision* e *Recall*) no conjunto de treino e no conjunto de teste e  $X$  ao valor do *F1-Score* no conjunto de teste.

Quando o valor de  $f$  é semelhante entre os modelos é preciso ter em conta outros fatores como o número de *features* usadas no treino do modelo e a simplicidade da sua implementação na aplicação. Deste modo, o processo de seleção aplicado considera:

1. Maior  $f$ .
2. Menor número de *features* usadas para treinar o modelo.
3. Maior simplicidade na implementação do modelo.

Os resultados finais obtidos para cada modelo estão descritos no capítulo seguinte.

## 4.2 Calibração da Cor

A versão anterior da aplicação apresentou problemas no âmbito da consistência da cor das imagens, devido à influência de fatores externos como diferentes tipos de iluminação e de dispositivos entre as medições. No entanto, a análise colorimétrica requer que as cores obtidas sejam o mais consistente possível independentemente das condições de captura das imagens.

De modo a atenuar este problema foram estudados cinco novos algoritmos, já descritos no capítulo do trabalho relacionado (Secção 3.2) e ainda uma versão melhorada do algoritmo usado no trabalho anterior descrita de seguida na secção 4.2.1. Estes algoritmos baseiam-se maioritariamente em operações de normalização da cor com recurso a *designs* específicos do sensor (marcas e molduras) e à introdução de cores de referência.

Para selecionar o algoritmo mais apropriado foram obtidas imagens que, após corrigidas com recurso aos algoritmos, foram comparadas entre si e com as referências de modo a escolher o melhor processo de correção disponível. Os resultados finais obtidos para cada modelo estão descritos no capítulo seguinte.

Os programas criados para testar os vários algoritmos foram desenvolvidos nas linguagens de programação *C++* e *Python* e recorrem à biblioteca *OpenCV* para todas as operações de processamento de imagem.

#### 4.2.1 Modificação do Algoritmo Usado na Versão Anterior

O algoritmo usado na versão anterior da aplicação apresentava problemas na consistência entre as imagens capturadas face a diferenças de iluminação ou de dispositivo. No entanto, nas conclusões do trabalho anterior foi proposta a possível melhoria dos resultados através da inclusão no sensor de cores de referência adicionais. O algoritmo seria idêntico ao descrito na secção 3.2.3, mas utilizando oito cores de referência em vez das quatro existentes nas matrizes aplicadas na fórmula 3.13.

De modo a incluir esta abordagem no conjunto de algoritmos a testar para uma possível introdução na aplicação, foi necessário criar um novo *design* do sensor que tivesse as referências necessárias.

Como não foi proposto nenhum conjunto de cores em particular, para além das quatro já presentes (branco, vermelho, azul e verde), foi usado como referência o *ColorChecker*, um alvo para calibração de câmaras fotográficas [51]. Das 24 cores disponíveis no alvo, procurou-se valores semelhantes aos quatro já existentes no sensor. Assim, foram escolhidos valores da zona das cores primárias (azul, verde, vermelho e amarelo) e da zona da escala de cinzentos, onde foi selecionado o branco e três tons de cinzento diferentes. As cores selecionadas foram depois adicionadas ao sensor criado através do *software Adobe Illustrator*. O *design* final está apresentado na figura 4.6, sendo os valores RGB das cores de referência descritos na tabela 4.8.

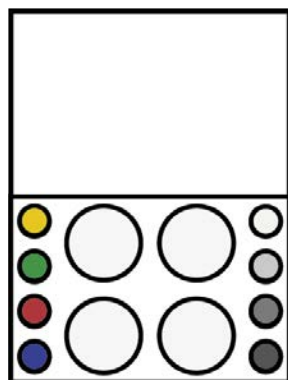


Figura 4.6: *Design* final do sensor com as oito cores de referência.

#### 4.2.2 Captura e Correção da Imagem

A maioria dos algoritmos depende de um determinado *design* do sensor para realizar as correções. Assim, foram desenhados sensores específicos para cada algoritmo com cores de três intensidades diferentes: sensor vazio e intensidade média e alta. Os sensores foram

	R	G	B
Amarelo	231	199	31
Verde	70	148	73
Vermelho	175	54	60
Azul	56	61	159
Branco	243	243	242
Cinza claro	200	200	200
Cinza médio	160	160	160
Cinza escuro	122	122	121

Tabela 4.8: Valores RGB das cores de referência do sensor

criados utilizando o *software Adobe Illustrator* e impressos tanto em papel de impressão normal como no papel de cromatografia *Whatman*, utilizado no CENIMAT para o fabrico dos biossensores.

Foi feita uma comparação inicial entre as cores obtidas com estas duas impressões e verificou-se a existência de uma grande diferença entre elas, principalmente na presença de uma fonte de luz forte (Figura 4.7). Concluiu-se então que o tipo de papel usado poderia influenciar os resultados obtidos e optou-se por utilizar o papel do laboratório em todos os testes realizados.

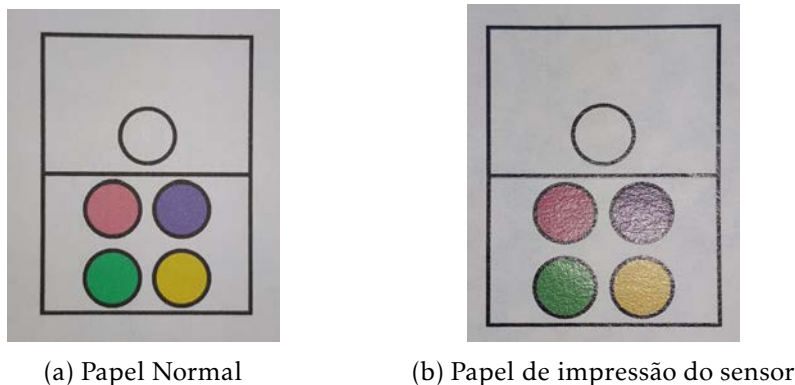


Figura 4.7: Comparação entre a impressão em papel normal e o papel usado na criação dos sensores, fotografados nas mesmas condições.

Para analisar a influência da luz e de diferentes dispositivos, os sensores impressos foram depois fotografados em três iluminações distintas: solar, fluorescente e LED, utilizando dois *smartphones* distintos: *Samsung S8* (gama alta) e *Motorola G3* (gama média).

As imagens obtidas foram depois corrigidas com recurso aos algoritmos e comparadas entre si e com as referências escolhidas. Como referência para as comparações foi utilizado o *scanner* do laboratório do CENIMAT (*Canon MG5250*), idêntico ao usado na obtenção do conjunto de dados para a análise colorimétrica. Outra referência aplicada foi o ficheiro criado no *Adobe Illustrator* onde foram criados os sensores.

As imagens originais e respetivas correções para cada algoritmo estão disponíveis no

apêndice B.

### 4.2.3 Comparação da Cor

Após a correção das imagens procede-se à extração da média da cor RGB de cada poço do sensor e das imagens de referência correspondentes. As cores são depois convertidas para o modelo Lab, necessário para o cálculo da métrica escolhida para a comparação das várias imagens.

A comparação baseia-se numa métrica de diferenciação da cor ( $\Delta$ ) que quantifica a diferença da percepção humana entre duas cores. Procura reproduzir a capacidade do olho humano para distinguir duas cores. Estas métricas são usadas predominantemente na indústria, para o fabrico de tintas e aparelhos de medição, por exemplo, e que permitem identificar diferenças de cor entre uma amostra e um determinado padrão, para que a cor possa ser considerada a mesma, independentemente do observador ou aparelho de medição usado [10]. Existem várias fórmulas para o cálculo deste valor com diferentes graus de precisão, baseadas no cálculo da distância euclidiana no modelo de cor Lab, com diversos fatores de modificação adicionais que procuram aproximar o resultado à diferença observada pelo olho humano [10].

Neste caso é aplicada a fórmula  $\Delta E_{2000}$ , o *standard* aceite atualmente pela Organização Internacional de Normalização (ISO), representada na equação 4.2 [40].

$$\Delta E_{2000} = \sqrt{\left(\frac{\Delta L'}{K_L S_L}\right)^2 + \left(\frac{\Delta C'}{K_C S_C}\right)^2 + \left(\frac{\Delta H'}{K_H S_H}\right)^2 + R_T \left(\frac{\Delta C'}{K_C S_C}\right) \left(\frac{\Delta H'}{K_H S_H}\right)} \quad (4.2)$$

Com recurso a esta fórmula são calculados três fatores para a comparação dos vários algoritmos:

- Soma dos  $\Delta E_{2000}$  calculados entre as cores das referências e as imagens corrigidas com os diferentes algoritmos. As referências são as imagens dos sensores obtidas com um *scanner* (Tabela A.3) e os sensores do ficheiro digital (Tabela A.4).
- Soma do  $\Delta E_{2000}$  calculados entre as imagens corrigidas obtidas com os dois modelos de *smartphone* analisados (Tabela A.1).
- Soma do  $\Delta E_{2000}$  calculados entre as imagens corrigidas obtidas com as várias iluminações analisadas para cada modelo de *smartphone* (Tabela A.2).

Para além destes fatores existem outras situações que influenciam a escolha do algoritmo a implementar, como a dificuldade da sua implementação na aplicação já existente e ainda a sua necessidade de recursos.

Os resultados finais obtidos para cada algoritmo estão descritos no capítulo seguinte.





## RESULTADOS

Neste capítulo são apresentados os resultados obtidos nos processos da análise colorimétrica e calibração já descritos no capítulo 4. Em relação à análise colorimétrica, são detalhados os resultados das métricas de avaliação da classificação para todos os algoritmos e combinações de *features* testados. No que diz respeito à calibração da cor são apresentados os resultados das métricas de comparação entre os vários algoritmos.

Para além da apresentação dos resultados é realizada uma reflexão sobre os mesmos de modo a determinar quais das opções estudadas são as mais indicadas para introdução na aplicação.

### 5.1 Análise Colorimétrica

No estudo da análise colorimétrica foram treinados múltiplos classificadores para cada uma das substâncias em análise, de modo a procurar obter o melhor resultado possível.

Os classificadores foram avaliados com recurso a um fator  $f$  definido na secção 4.1.2.3. Este fator procura evitar o *overfitting*, tendo em conta não só o valor mais elevado do  $F_1$  Score mas também a menor diferença entre os resultados das várias métricas obtidas com *K-Fold Cross Validation* e na validação com o conjunto de teste.

Para estabelecer um limite inferior para os resultados da classificação foi treinado um classificador aleatório para cada um dos modelos de dados. Criado a partir do método *DummyClassifier* [32], o classificador faz previsões aleatórias e, consequentemente, obtém resultados fracos para todas as métricas. Assim, constitui uma boa comparação para o desempenho mínimo que os algoritmos testados devem atingir. Os resultados para este classificador estão presentes na tabela 5.1.

Modelos de Dados		F1 Score	Recall	Precision
Concentrações de glucose		0.08	0.10	0.09
Quatro níveis de glucose	Jejum	0.27	0.32	0.28
	Duas horas	0.28	0.38	0.31
Cinco níveis de glucose	Jejum	0.22	0.23	0.23
	Duas horas	0.22	0.26	0.24
Concentrações de ácido úrico		0.16	0.16	0.16
Níveis de ácido úrico		0.40	0.41	0.41

Tabela 5.1: Resultados do classificador aleatório para os vários modelos de dados analisados.

### 5.1.1 Classificação da Glucose

Na análise da glucose foram treinados classificadores para três modelos de dados:

1. **Dez concentrações fornecidas:** 0, 1.25, 1.5, 2, 5, 7, 10, 12.5, 15, 20 (mM).
2. **Quatro níveis de referência:** vazio, hipoglicemia, normal/pré-diabetes e diabetes.
3. **Cinco níveis de referência:** vazio, hipoglicemia, normal, pré-diabetes e diabetes.

Os dois primeiros modelos foram estudados no trabalho anterior [37], sendo o segundo criado a partir dos níveis de referência da diabetes, devido aos resultados insuficientes obtidos na classificação das dez concentrações. Estes dividem-se ainda em estado de jejum ou duas horas após a refeição, pois estes fatores influenciam o significado de determinada concentração de glucose no indivíduo.

Como foi possível aumentar o número de repetições em algumas das classes optou-se também por incluir um modelo adicional baseado na separação total dos níveis de referência da glucose [34]. A implementação deste modelo permitiria oferecer ao utilizador um resultado mais detalhado sobre o seu estado, o que seria vantajoso.

Para cada um destes modelos, a classificação foi testada com todas as combinações de *features* descritas na secção 4.1.1.3. As cardinalidades dos vários conjuntos de dados referidos estão referidas na secção 4.1.1.1.

#### 5.1.1.1 Dez Concentrações Fornecidas

Os resultados obtidos para as concentrações foram na sua maioria perto dos 0.5 de  $F_1$  Score. Este resultado, apesar de superior a uma classificação aleatória (tabela 5.1), continua a ser muito insuficiente visto que corresponde a valores de precisão e sensibilidade de cerca de 50%, o que não é aceitável para um dispositivo de diagnóstico.

O modelo SVM baseado nos três momentos de cor em YUV (sem PCA), obteve resultados por volta dos 0.7. Para além de superar os restantes modelos em análise, este

resultado é bastante melhor que os obtidos para as concentrações no trabalho anterior, que rondavam também os 0.5 [37].

Os resultados integrais estão disponíveis no apêndice D e um resumo dos melhores resultados pode ser consultado na tabela 5.2.

Ambos os modelos que envolvem os três momentos de cor em YUV obtiveram valores semelhantes de  $f$ . No entanto, no espectro global da análise os resultados são muito reduzidos quando comparados com os dos dois outros modelos de dados, pelo que nenhum destes foi selecionado para implementação na aplicação.

Classificador	PCA	Modelo	Conjunto de Teste			Cross-Validation			$f$
			F1	R	P	F1	R	P	
KNN	Não	3 momentos YUV	0.65	0.66	0.65	0.64	0.66	0.66	0.63
	Sim	Média S e V (HSV)	0.53	0.52	0.54	0.49	0.50	0.50	0.43
SVM	Não	3 momentos YUV	0.71	0.72	0.73	0.69	0.69	0.73	0.65
	Sim	Média YUV	0.50	0.49	0.53	0.50	0.52	0.52	0.46
Random Forest	Não	Média L e S (HLS)	0.44	0.46	0.44	0.43	0.46	0.42	0.41
	Sim	Média S e V (HSV)	0.46	0.49	0.46	0.41	0.45	0.42	0.33

Tabela 5.2: Resultados dos melhores modelos para a classificação das concentrações da glucose (F1 -  $F_1 Score$ , R - Recall, P- Precision).

#### 5.1.1.2 Quatro Níveis

Os resultados obtidos na análise dos quatro níveis de referência foram muito superiores aos verificados na análise das concentrações e dos cinco níveis, rondando os 0.85 de  $F_1 Score$ . No entanto, estes são inferiores quando comparados com os resultados dos quatro níveis obtidos no trabalho anterior [37], que rondavam os 0.95 nos melhores resultados.

Em relação ao modelo do estado de jejum o resultado mais elevado chega aos 0.84 de  $F_1 Score$  para o algoritmo KNN. Este ocorre, tal como na análise das concentrações, quando são aplicados os três momentos da cor no espaço de cor YUV (sem PCA).

Quanto ao modelo das duas horas após a refeição, os resultados mais elevados são muito semelhantes entre os três algoritmos, rondando os 0.85 de  $F_1 Score$ . Tanto no KNN como no SVM, os melhores resultados ocorrem nos modelos mais simples, isto é, aqueles onde são usadas apenas duas *features*. No entanto, o resultado obtido para o classificador KNN ao usar a média das componentes L e S do espaço HLS é ligeiramente mais elevado. O algoritmo *Random Forest* apresenta resultados igualmente favoráveis, mas aplica um maior número de *features*.

Os resultados integrais estão disponíveis no apêndice D e um resumo dos melhores resultados pode ser consultado na tabela 5.3, para o jejum e na tabela 5.4 para duas horas após a refeição.

Classificador	PCA	Modelo	Conjunto de Teste			Cross-Validation			f
			F1	R	P	F1	R	P	
KNN	Não	3 momentos YUV	0.84	0.85	0.84	0.84	0.84	0.85	0.82
	Sim	Média LAB	0.76	0.76	0.76	0.77	0.77	0.79	0.71
SVM	Não	Média XYZ	0.78	0.78	0.78	0.77	0.77	0.79	0.76
	Sim	Média HLS	0.75	0.74	0.77	0.72	0.73	0.76	0.69
Random Forest	Não	Média S e V (HSV)	0.74	0.73	0.75	0.74	0.74	0.75	0.72
	Sim	3 momentos YUV	0.80	0.80	0.80	0.80	0.81	0.82	0.78

Tabela 5.3: Resultados dos melhores modelos para a classificação dos quatro níveis em estado de jejum (F1 -  $F_1 Score$ , R - Recall, P- Precision).

Classificador	PCA	Modelo	Conjunto de Teste			Cross-Validation			f
			F1	R	P	F1	R	P	
KNN	Não	Média L e S (HLS)	0.86	0.86	0.87	0.86	0.85	0.89	0.83
	Sim	Média L e S (HLS)	0.86	0.86	0.87	0.86	0.85	0.89	0.83
SVM	Não	Média S e V (HSV)	0.85	0.85	0.85	0.86	0.85	0.87	0.82
	Sim	Média S e V (HSV)	0.85	0.85	0.85	0.86	0.85	0.87	0.82
Random Forest	Não	3 momentos YUV	0.86	0.88	0.85	0.86	0.86	0.86	0.84
	Sim	Média HLS	0.85	0.87	0.83	0.85	0.86	0.84	0.83

Tabela 5.4: Resultados dos melhores modelos para a classificação dos quatro níveis, duas horas após a refeição (F1 -  $F_1 Score$ , R - Recall, P- Precision).

### 5.1.1.3 Cinco Níveis

Apesar de superiores aos das concentrações, os resultados obtidos na análise dos cinco níveis de referência foram inferiores aos verificados apenas com os quatro níveis, ficando abaixo dos 0.8 de  $F_1 Score$ .

Em relação ao modelo de jejum o melhor resultado em termos de  $F_1 Score$  atinge os 0.77, sendo o modelo novamente baseado novamente nos três momentos de cor em YUV (sem PCA). No entanto, o seu valor do factor  $f$  é mais reduzido devido às diferenças entre a *cross-validation* e o conjunto de teste. Assim, o melhor resultado para este modelo ocorre com o algoritmo SVM baseado na média no espaço de cor XYZ.

Quanto ao modelo das duas horas após a refeição, os resultados mais elevados ocorrem com o algoritmo KNN baseado nos três momentos de cor no espaço HSV (sem PCA). O algoritmo *Random Forest* apresenta resultados semelhantes mas é baseado no espaço de cor Lab que, em comparação com o HSV, requer uma conversão mais complexa a partir do valor RGB extraído inicialmente.

Os resultados integrais estão disponíveis no apêndice D e um resumo dos melhores resultados pode ser consultado na tabela 5.5, para o jejum e na tabela 5.6 para duas horas após a refeição.

Sendo este modelo de dados baseado nos níveis de referência da diabetes, era preferível a sua implementação na aplicação, pois poderia fornecer ao utilizador informações mais úteis sobre o seu estado de saúde. No entanto, apesar dos resultados obtidos terem sido promissores, os valores continuam abaixo dos verificados com os quatro níveis, não justificando a sua inclusão na aplicação.

Classificador	PCA	Modelo	Conjunto de Teste			Cross Validation			f
			F1	R	P	F1	R	P	
KNN	Não	3 momentos YUV	0.77	0.76	0.78	0.79	0.79	0.80	0.70
	Sim	Média L e S (HLS)	0.67	0.67	0.67	0.66	0.66	0.68	0.65
SVM	Não	Média XYZ	0.72	0.72	0.72	0.72	0.72	0.73	0.71
	Sim	Média L e S (HLS)	0.66	0.65	0.68	0.68	0.68	0.69	0.60
Random Forest	Não	Média HLS	0.69	0.69	0.71	0.68	0.68	0.70	0.66
	Sim	Média S e V (HSV)	0.66	0.66	0.67	0.66	0.66	0.67	0.65

Tabela 5.5: Resultados dos melhores modelos para a classificação dos cinco níveis em estado de jejum (F1 -  $F_1$  Score, R - Recall, P- Precision).

Classificador	PCA	Modelo	Conjunto de Teste			Cross Validation			f
			F1	R	P	F1	R	P	
KNN	Não	3 momentos HSV	0.79	0.79	0.79	0.78	0.78	0.80	0.76
	Sim	Média HSV	0.70	0.68	0.75	0.70	0.69	0.75	0.69
SVM	Não	Média LAB	0.76	0.74	0.78	0.76	0.76	0.78	0.73
	Sim	Média YUV	0.74	0.73	0.75	0.74	0.73	0.77	0.71
Random Forest	Não	3 momentos LAB	0.78	0.78	0.79	0.77	0.77	0.78	0.75
	Sim	Média LAB	0.75	0.74	0.78	0.73	0.72	0.78	0.71

Tabela 5.6: Resultados dos melhores modelos para a classificação dos cinco níveis, duas horas após a refeição (F1 -  $F_1$  Score, R - Recall, P- Precision).

#### 5.1.1.4 Conclusões

**Dez concentrações** Em semelhança ao observado no trabalho anterior, a classificação das dez concentrações manteve os resultados baixos na ordem dos 0.5 de  $F_1$  Score. No entanto, ao utilizar os três momentos da cor no espaço YUV (sem PCA), tanto no classificador KNN como no SVM, os resultados atingiram valores superiores entre os 0.65 a 0.75. Isto indica que este conjunto particular de *features* pode trazer vantagens para a classificação deste tipo de dados, quando comparado com outras configurações.

Ao aumentar o número de amostras das classes anteriores de 24 para mais de cem exemplos, era de esperar que houvesse um aumento geral dos resultados. No entanto, não foi possível usar o conjunto de dados na totalidade, o que levou a um grande desequilíbrio face às concentrações não aumentadas. Esta diferença pode ter limitado a capacidade dos

classificadores para obter um resultado superior. Independentemente das *features* usadas, podemos ainda observar que para os classificadores KNN e SVM, os resultados obtidos sem a aplicação de PCA foram em média 10% superiores àqueles onde foi feita a redução de dimensionalidade.

Apesar do aumento do tamanho do conjunto de dados face ao do trabalho anterior, a classificação das concentrações apresentou resultados insuficientes que não permitem a sua inclusão na aplicação.

**Quatro Níveis** De entre todas as disposições de dados estudadas, o modelo dos quatro níveis obteve os melhores resultados, tanto para o estado de jejum como para o de duas horas após a refeição. Deste modo, tal como no trabalho anterior, foi este o modelo de dados implementado na aplicação.

No caso do estado de jejum, na melhor situação são atingidos os 0.84 de  $F_1 Score$  através do algoritmo KNN baseado nos três momentos de cor em YUV. No entanto, no trabalho anterior, os modelos testados para o estado de jejum ultrapassavam os 0.9 de  $F_1 Score$ , sendo que o ótimo atingia os 0.94. Apesar do aumento do conjunto de dados, nenhum dos modelos novos conseguiu atingir estes valores. Na versão anterior já se verificavam um algumas diferenças entre as cardinalidades das várias classes. Os número de exemplos adicionados veio acentuar essas diferenças, levando a uma diferença de 184 exemplos entre a classe menos representada (vazio) e a classe mais representada (Normal/Pré-Diabetes).

Na situação das duas horas após a refeição, foi possível obter perto dos 0.85 de  $F_1 Score$  com os diferentes algoritmos. No entanto, o classificador KNN baseado na média das componentes L e S obteve um melhor resultado baseado em menos *features* que os outros modelos. No trabalho anterior, os modelos testados rondavam os 0.80 de  $F_1 Score$ , sendo que o ótimo atingiu os 0.91. Existia também um equilíbrio entre as três classes não vazias. Com o aumento de exemplos nas duas classes intermédias surgiu um desequilíbrio acentuado entre estas e as classes dos extremos. Apesar do aumento do conjunto de dados, nenhum dos novos modelos testados conseguiu atingir esse valor mais elevado. No entanto, os valores obtidos mantiveram-se perto dos 0.80 atingidos anteriormente. Aliás, ao comparar todos os modelos testados no trabalho anterior com os novos modelos baseados nas mesmas *features*, podemos constatar que os modelos novos obtêm melhores resultados em cerca de 0.05 de  $F_1 Score$ , independentemente do algoritmo usado.

**Cinco Níveis** Por último, o modelo dividido em cinco níveis fornece uma informação, ainda que qualitativa, mais útil para o diagnóstico do utilizador. Os resultados obtidos na ordem dos 0.7 a 0.8 de  $F_1 Score$ , tanto no estado de jejum como duas horas após a refeição, mostram potencial para a futura implementação deste modelo de dados.

Apesar de haver maior equilíbrio entre o número de elementos das classes do que no modelo dos quatro níveis, os resultados revelaram-se inferiores. Isto indica que o número de exemplos existentes por cada classe não é suficiente para obter melhores resultados. É

por isso necessário aumentar a cardinalidade do conjunto de dados, mantendo o equilíbrio entre as mesmas.

Como os resultados foram inferiores aos fornecidos pelo modelo dos quatro níveis optou-se por não introduzir este modelo na aplicação.

### 5.1.2 Classificação do Ácido Úrico

Na análise do ácido úrico foram treinados classificadores para dois modelos de dados:

1. **Oito concentrações fornecidas:** 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1 (mM).
2. **Três níveis de referência:** baixo (até 0.18 mM), normal (0.18 mM a 0.4 mM) e alto (superior a 0.4 mM).

Inicialmente pretendia-se estudar apenas a classificação para as oito concentrações fornecidas. No entanto, no decorrer da análise verificou-se que os resultados dessa classificação não seriam elevados o suficiente para justificar a sua introdução na aplicação móvel. Assim, foi criado um segundo modelo semelhante aos criados para a glucose, mas baseado em níveis de referência do ácido úrico, sugeridos pelo autor do sensor [33].

Os níveis de referência do ácido úrico divergem consoante o sexo do indivíduo. Idealmente deviam ser analisados níveis de ambos os sexos, de modo a fornecer uma informação completa a todos os utilizadores. No entanto, como se trata da primeira análise deste conjunto de dados, a prioridade é identificar se existe potencial para a classificação desta substância com a gama de cores existente. O elevado nível de ácido úrico está fortemente associado a uma doença inflamatória chamada gota, que é mais prevalente nos homens [42]. Assim, optamos por analisar apenas os níveis do sexo masculino.

Para cada modelo de dados foram testadas todas as combinações de *features* referidas na secção 4.1.1.3. As cardinalidades dos vários conjuntos de dados referidos estão detalhadas na secção 4.1.1.

#### 5.1.2.1 Oito Concentrações Fornecidas

Tal como sucedeu com a glucose, os resultados obtidos para as concentrações foram na sua maioria perto dos 0.5 de  $F_1$  Score. Este resultado, apesar de superior a uma classificação aleatória (tabela 5.1), continua a ser muito insuficiente visto que corresponde a valores de precisão e sensibilidade de abaixo de 50%, o que não é aceitável para um dispositivo de diagnóstico.

Os resultados integrais estão disponíveis no apêndice C e um resumo dos melhores resultados pode ser consultado na tabela 5.7. Independentemente da combinação de *features* aplicada, os resultados obtidos para este modelo de dados foram claramente insuficientes para poder fornecer qualquer informação fiável ao utilizador.

Classificador	PCA	Modelo	Conjunto de Teste			Cross Validation			f
			F1	R	P	F1	R	P	
KNN	Não	Média H e S (HSV) e H e S (HLS)	0.46	0.46	0.47	0.42	0.45	0.45	0.39
	Sim	Média HSV	0.38	0.39	0.41	0.38	0.40	0.42	0.37
SVM	Não	Média HSV	0.51	0.51	0.51	0.46	0.48	0.50	0.42
	Sim	Média H e S (HSV)	0.47	0.49	0.49	0.42	0.44	0.47	0.36
Random Forest	Não	Média H e S (HSV)	0.40	0.41	0.42	0.37	0.40	0.39	0.33
	Sim	Média H e S (HSV)	0.41	0.43	0.44	0.39	0.42	0.42	0.37

Tabela 5.7: Resultados dos melhores modelos para a classificação das concentrações do ácido úrico (F1 -  $F_1 Score$ , R - Recall, P- Precision).

### 5.1.2.2 Três Níveis

Dado que os resultados obtidos para as oito concentrações não foram favoráveis foi necessário criar um modelo baseado nos níveis de referência do ácido úrico.

Independentemente do algoritmo aplicado, os valores de  $F_1 Score$  obtidos foram semelhantes entre si e muito superiores aos verificados com as concentrações, rondando os 0.80. No entanto, o classificador SVM baseado na média das componentes H e S dos espaços HSV e HLS, obteve um melhor resultado em termos do factor  $f$ .

Os resultados integrais estão disponíveis no apêndice C e um resumo dos melhores resultados pode ser consultado na tabela 5.8.

Classificador	PCA	Modelo	Conjunto de Teste			Cross Validation			f
			F1	R	P	F1	R	P	
KNN	Não	Média RGB	0.78	0.79	0.79	0.78	0.78	0.79	0.77
	Sim	Média HLS	0.81	0.81	0.82	0.80	0.80	0.81	0.78
SVM	Não	Média HSV	0.80	0.80	0.81	0.80	0.80	0.82	0.77
	Sim	Média H e S (HSV) e H e S (HLS)	0.81	0.82	0.81	0.81	0.81	0.83	0.79
Random Forest	Não	Média LAB	0.80	0.81	0.82	0.80	0.80	0.81	0.78
	Sim	Média LAB	0.80	0.80	0.80	0.81	0.81	0.82	0.75

Tabela 5.8: Resultados dos melhores modelos para a classificação dos três níveis do ácido úrico (F1 -  $F_1 Score$ , R - Recall, P- Precision).

### 5.1.2.3 Conclusões

**Concentrações** Em semelhança ao observado com as concentrações da glucose, a classificação das oito concentrações do ácido úrico não obteve resultados favoráveis. Independentemente do algoritmo e combinação de *features* testados, os valores atingidos foram muito fracos, abaixo dos 0.5 de  $F_1 Score$ . Assim, apesar desta informação quantitativa ser



mais relevante para o diagnóstico do utilizador, os resultados desta ordem não permitem a sua inclusão na aplicação.

Ao observar o conjunto de dados, podemos constatar que há equilíbrio total nas várias classes. No entanto, o número de exemplos para cada classe é muito reduzido (48 exemplos), semelhante ao que acontecia no trabalho anterior com a glucose. Com o aumento do número de amostras de algumas classes da glucose (de 24 para 120), as suas concentrações conseguiram atingir um valor na ordem dos 0.70. Seria então interessante aumentar o número de exemplos das várias concentrações do ácido úrico, para verificar se será possível atingir também resultados dessa ordem.

**Três Níveis** Através do modelo de dados dividido em três níveis, conseguimos atingir resultados muito superiores, na ordem dos 0.80 de  $F_1$  Score.

Neste caso, foi possível obter resultados perto dos 0.81 de  $F_1$  Score com os diferentes algoritmos. No entanto, o classificador SVM baseado na média das coordenadas H e S da cor nos espaços HSV e HLS, obteve um melhor valor no fator  $f$  do que os restantes.

Apesar de não fornecer um resultado quantitativo sobre a substância, este modelo permite obter resultados satisfatórios fornecendo ao utilizador uma informação qualitativa relevante sobre o seu nível de ácido úrico.

Ao observar o conjunto de dados, constatamos que há equilíbrio entre as classes mais elevadas, sendo que a mais baixa tem uma diferença de 48 elementos. Seria então importante aumentar também o número de elementos desta classe para manter o equilíbrio do conjunto, o que poderia levar a melhores resultados na classificação.

**Conjunto de Dados** Geralmente, o conjunto de dados do ácido úrico fornecido pelo CENIMAT não tem o tamanho suficiente para facultar um melhor resultado na classificação. No entanto, os resultados obtidos com esta gama de cor através dos níveis de referência mostram que existe potencial para obter bons resultados na classificação desta gama de cores. Independentemente do modelo de dados (concentrações ou níveis de referência), para alcançar melhores resultados na classificação, propõe-se o aumento do número de exemplos para cada classe, mantendo o equilíbrio entre as mesmas.

Outra lacuna na classificação provém do facto de o conjunto de dados não ter exemplos para o estado em que o sensor está vazio (sem reagente). Isto leva a resultados improváveis caso o utilizador teste um sensor vazio. Seria então também importante incluir um número de amostras vazias semelhante às restantes classes para que seja possível fazer esta classificação.

## 5.2 Calibração

No estudo da calibração foram implementados e testados múltiplos algoritmos descritos no estado da arte (Secção 3.2). Estes procuram melhorar o desempenho do algoritmo já

existente na aplicação, eliminando a influência da iluminação nas imagens capturadas, de modo a reduzir o seu impacto na análise colorimétrica.

Para selecionar o melhor método de calibração, os algoritmos foram comparados usando a métrica  $\Delta E_{2000}$  [40] que descreve a diferença entre duas cores, através do processo já detalhado na secção 4.2.3. A partir dos resultados dessa comparação foi selecionado um dos algoritmos para introdução na aplicação.

### 5.2.1 Seleção do Algoritmo a Implementar

Os sete algoritmos estudados foram comparados com recurso a três fatores:

1. Diferença entre as cores da imagem e das duas referências.
2. Diferença entre as cores de imagens obtidas com os diferentes modelos de *smartphone*.
3. Diferença entre as cores de imagens obtidas com as diferentes iluminações.

Tal como uma das referências usadas, os classificadores são treinados a partir de imagens de amostras obtidas com o scanner do CENIMAT. Como pretendemos aproximar as características das imagens capturadas no dispositivo móvel das imagens usadas para o treino dos classificadores, precisamos de valorizar a menor diferença entre elas.

Ao procurar reduzir a influência de fatores externos na imagem, podemos estar a adulterar a cor que pretendemos corrigir. Esta alteração da cor é particularmente nociva, visto que não será muito relevante manter a consistência entre dispositivos e iluminações diversas, se a cor for completamente diferente da presente no conjunto de dados fornecido para o treino. Deste modo, o fator mais importante na comparação é a diferença entre as cores das imagens após a correção e as suas referências. Caso o algoritmo apresente uma diferença de cores aceitável podemos então observar a diferença entre os fatores externos: diferentes modelos de *smartphone* e diferentes iluminações.

Os resultados obtidos para a comparação dos vários algoritmos estão detalhados na tabela 5.9.

Todos os algoritmos analisados obtiveram melhores resultados que o existente na versão anterior da aplicação. O valor de  $\Delta E_{2000}$  mais reduzido foi obtido com a versão melhorada do algoritmo anterior (secção 4.2.1), seguido de perto pelos restantes algoritmos. No entanto, o algoritmo proposto por Zhang et al. e o algoritmo anterior atingiram resultados muito mais elevados. Estes resultados elevados indicam logo à partida que estes métodos não são os mais apropriados para a correção que pretendemos.

O algoritmo anterior [37] apresenta resultados aceitáveis para a diferença entre os modelos de *smartphone*. Porém, atinge um dos piores resultados em relação à comparação com as referências e entre iluminações.

A proposta de Zhang et al.[54] é a mais complexa em análise, no que diz respeito à aplicação de transformações entre modelos de cor menos triviais e de sucessivas regressões polinomiais para correção da imagem. No entanto, obteve o pior resultado de todos

	Anterior	Modificado	Jia et al.	Lee et al.	Lopez Ruiz et al.	Zhang et al.	Souza et al.
Comparação entre cores	3134.30	1329.72	1285.20	1613.77	1688.19	2263.75	1365.79
Comparação entre modelos	266.23	157.31	243.75	149.43	192.06	374.64	445.92
Comparação entre luzes	694.67	543.47	617.72	474.98	447.82	1013.16	430.51
Total	4095.21	2030.50	2146.66	2238.18	2328.06	3651.55	2242.22

Tabela 5.9: Somas do  $\Delta E_{2000}$  para a comparação dos algoritmos estudados.

os novos algoritmos em estudo, sendo ainda o pior de todos em relação à comparação da cor entre diferentes iluminações.

O algoritmo modificado apresenta resultados favoráveis em todos os fatores, sendo um dos melhores em termos da diferença entre as cores e entre diferentes dispositivos. Outro ponto positivo é o facto deste aplicar uma única transformação na imagem, baseada apenas nas cores de referência. Para além da simplicidade do método, a sua introdução na aplicação também é relativamente simples, visto que se trata do algoritmo já implementado, modificado apenas para incluir um maior número de cores de referência.

Outras abordagens propostas por Jia et al.[14] e Lee e Woo[19] apresentam resultados favoráveis em todas as vertentes, sendo os melhores em relação à diferença entre os modelos de *smartphone* e à diferença entre as cores, respetivamente. No entanto, ambos incorrem em problemas semelhantes, que dificultam a sua implementação no nosso sistema. Estes métodos requerem a extração da cor de várias zonas específicas dos sensores (molduras e círculos) que, para serem recolhidas, envolvem múltiplas operações de processamento de imagem. Estas operações tornam o seu processo menos eficiente que a maioria dos algoritmos estudados. Estes métodos requerem ainda uma modificação total do *design* do biossensor. Como o sensor está sujeito a uma disposição restrita devido aos canais de microfluidica, quaisquer modificações mais complexas têm de ser estudadas e aprovadas pelo CENIMAT.

Como esperado, o algoritmo de Lopez-Ruiz et al.[20] atingiu bons resultados em termos da diferença entre iluminações e entre dispositivos. No entanto, apresentou resultados menos favoráveis na comparação entre as cores das imagens e das referências. Tal deve-se ao facto de recorrer à lanterna do *smartphone* para tentar reduzir a influência da luz externa. Isso torna as cores mais consistentes entre os dispositivos com lanternas semelhantes e entre as diferentes iluminações, mas altera-as significativamente quando comparadas com as das referências. Para além disso, neste contexto foram testados dois dispositivos com lanternas semelhantes, mas caso sejam utilizados modelos com lanternas de intensidade ou cor diferentes, a diferença entre as luzes pode vir a prejudicar a

consistência entre os modelos de *smartphone*.

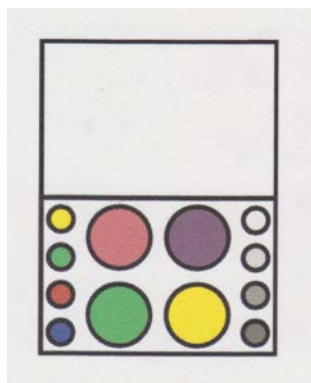
Por último, a proposta de Souza et al.[44] baseada numa cor relativa apresentou resultados favoráveis em termos da diferença entre cores e entre iluminações, sendo a melhor em relação às iluminações. No entanto, apresenta o pior resultado de todos no que diz respeito à diferença entre modelos de *smartphone*. Ademais, a sua implementação no sistema seria mais difícil. Como não é baseado na cor da imagem num determinado espaço de cor, mas sim em valores calculados a partir da cor da imagem e de uma referência, os valores resultantes não seriam em nada semelhantes às cores extraídas diretamente das placas para treino dos classificadores. Isto levaria à necessidade da alteração não só da parte referente à correção da cor mas também da extração e processamento de dados necessários para o treino dos classificadores.

Assim, baseado nos resultados obtidos e em todos os fatores adicionais foi escolhida para implementação a versão melhorada do algoritmo anterior, descrito na secção 4.2.1.

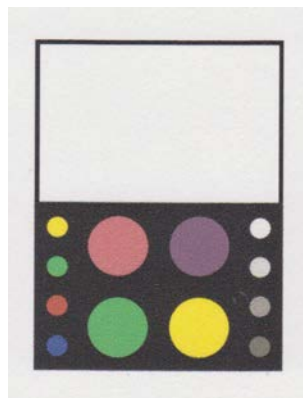
### 5.2.2 Alterações no Biossensor

Após a seleção do método de calibração, foi proposto ao CENIMAT o *design* final do sensor, presente na figura 5.1a. A partir do *design* proposto, o laboratório realizou algumas modificações no sensor para assegurar que o comportamento dos reagentes envolvidos na análise se mantinha consistente com o anterior.

As alterações consistem na introdução de texto na parte superior e na mudança de cor do fundo de branco para preto, originando um sensor semelhante ao presente na figura 5.1b.



(a) Sensor com fundo branco criado para testar o algoritmo de calibração.



(b) Sensor final com fundo preto.

Figura 5.1: Diferenças entre os sensores desenvolvidos para aplicar o novo método de calibração.

Face a esta mudança de cor optou-se por validar o método de correção seleccionado, comparando os resultados da correção para o sensor branco e para o sensor preto. Assim, foram fotografados os dois tipos de sensores sob as diferentes iluminações e com as

diferentes câmaras, tendo sido aplicado mesmo processo de teste usado na secção anterior. Os resultados obtidos estão descritos na tabela 5.10.

	Sensor Branco	Sensor Preto
Comparação entre Cores	1084.82	1117.34
Comparação entre Modelos	114.36	127.80
Comparação entre Luzes	312.32	335.76
Total	1511.51	1580.89

Tabela 5.10: Somas do  $\Delta E_{2000}$  para a comparação da correção entre as diferentes cores do sensor

Apesar de haver uma ligeira diferença entre os valores obtidos para os dois sensores, esta não é significativa o suficiente para se considerar que o método de correção escolhido é afetado pela mudança de cor do fundo do sensor. Podemos então considerar que o algoritmo selecionado previamente continua a ser a melhor escolha para a aplicação.



## IMPLEMENTAÇÃO

Neste capítulo é feita a descrição da implementação dos processos de deteção, calibração e análise colorimétrica seleccionados a partir dos resultados apresentados no capítulo anterior, na aplicação *Android*.

No seguimento do projeto anterior, a aplicação móvel continua a ser desenvolvida para o sistema *Android*<sup>1</sup>, com interface gráfica baseada no *Material Design*<sup>2</sup>, um conjunto de indicações para o design das aplicações *Android* fornecidas pela Google.

No que diz respeito à *interface* gráfica foram feitas poucas alterações, nomeadamente as necessárias à apresentação dos resultados do ácido úrico e à deteção. O conjunto de ecrãs principais da aplicação móvel está disponível no apêndice H.

Os algoritmos e métodos implementados para a deteção, extração e análise das cores do biossensor foram desenvolvidos em C++, utilizando as funcionalidades da biblioteca de processamento de imagem *OpenCV* [4] na sua versão destinada a desenvolvimento em *Android*. Esta biblioteca é *open source* e contém inúmeros algoritmos de visão computacional optimizados, sendo amplamente usada em contexto empresarial e de investigação. Inclui ainda funcionalidades de aprendizagem automática, que foram utilizadas para implementar na aplicação os classificadores seleccionados.

### 6.1 Deteção do Sensor

Um dos objetivos desta dissertação passa pela preparação do sistema anterior para detetar o novo biossensor proposto por Pinheiro [33].

Como podemos observar na figura 6.1b, o novo sensor difere do detetado anteriormente, presente na figura 6.1a, ao apresentar quatro zonas para análise em vez de apenas

---

<sup>1</sup><https://www.android.com>

<sup>2</sup><https://material.io/>

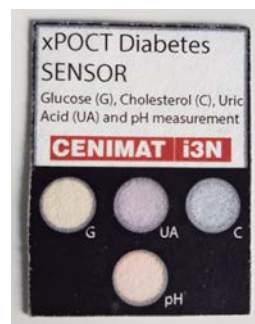
uma. Para além disso não estão presentes as quatro cores de referência definidas previamente para a calibração da cor da imagem.

Este novo dispositivo é um biossensor *multiplex* capaz de medir concentrações de glucose, ácido úrico, colesterol e ainda valores de pH. A cada um desses marcadores corresponde um círculo no sensor, denominado de poço. Nestes poços encontramos a cor que será extraída para fazer a análise, sendo esta obtida através de reações decorrentes da colocação da amostra química num único ponto, localizado na parte de trás do sensor. A amostra dirige-se depois para os diferentes poços através de canais de microfluídica [33].

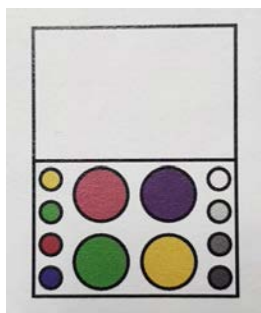
Para permitir a introdução do novo método de calibração da cor selecionado no capítulo anterior (Secção 5.2), foram incluídas no design do sensor oito zonas de referência de diferentes cores. Esta alteração dá origem ao sensor final presente na figura 6.1d.



(a) Biossensor usado anteriormente [37] com um poço para análise da glucose e quatro áreas de referência.



(b) Novo biossensor [33] com quatro poços para análise: glucose(G), ácido úrico (uA), colesterol (C) e pH.



(c) Modificação proposta para incluir o novo método de calibração.



(d) Biossensor final com a inclusão das zonas de referência propostas para a calibração.

Figura 6.1: Biossensores utilizados nas várias versões do projeto.

O processo de deteção segue o mesmo processo realizado anteriormente [37], com a alteração de fatores relacionados com as novas dimensões do sensor. As operações de processamento de imagem aplicadas são detalhadas na secção 2.1.2. O utilizador deve posicionar o sensor no interior das linhas mostradas no ecrã da aplicação. Caso a imagem obtida se encontre focada, será feita uma captura periódica de imagens dessa área de modo a localizar os poços onde se encontram as amostras que se pretende analisar. A



componente de detecção envolve as seguintes fases:

- **Verificar se a imagem está focada** - A imagem capturada é convertida para tons de cinzento e é aplicado um filtro de Laplace para detecção de contornos [5]. De seguida é calculada a variância dos *pixels* da imagem e o valor resultante é comparado com um limiar. Este limiar foi definido através do cálculo da variância para imagens focadas e desfocadas do sensor sob diferentes iluminações. Caso a variância da imagem seja inferior ao limiar estabelecido a imagem é considerada como desfocada.



(a) Imagem desfocada com variância = 7.



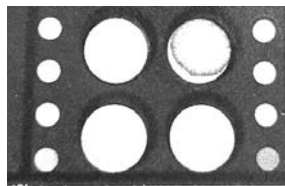
(b) Imagem focada com variância = 396.

Figura 6.2: Exemplo da variância para um sensor focado (b) e desfocado (a).

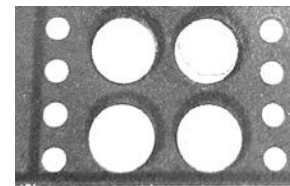
- **Detetar as doze zonas do sensor** - Caso a imagem esteja focada prossegue-se com a detecção do sensor. A zona onde se encontram os círculos é seleccionada com base nas dimensões do sensor. A partir desta zona, segue-se um conjunto de passos para identificar os círculos:
  1. **Converter a imagem para tons de cinzento** - O valor de cinzento aplicado a cada *pixel* corresponde ao valor máximo entre os seus componentes RGB.
  2. **Aumentar o brilho da imagem** - Os componentes RGB de cada pixel são multiplicados por um fator  $r$ , nunca ultrapassando 255. Este fator inicia a um mas é incrementado cada vez que se aplica o algoritmo. Quando  $r$  for superior a cinco volta a ser reposto a um.



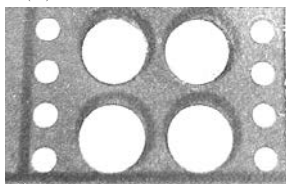
(a)  $r = 1$ .



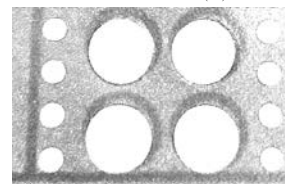
(b)  $r = 2$ .



(c)  $r = 3$ .



(d)  $r = 4$ .



(e)  $r = 5$ .

Figura 6.3: Exemplo dos efeitos do aumento do brilho no sensor.

3. **Aplicar um filtro Gaussiano** - Um dos filtros mais comuns para redução de ruído, onde é aplicado à imagem uma máscara baseada numa função Gaussiana [5].

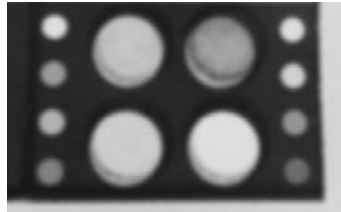


Figura 6.4: Resultado do filtro gaussiano.

4. **Binarização com *threshold* adaptativo** - Na binarização com *threshold* simples caso o *pixel* tenha um valor superior ao *threshold* é lhe atribuído 255, caso contrário fica a zero. No método adaptativo, o algoritmo calcula o *threshold* a aplicar. Neste caso, para cada *pixel*, o *threshold* é a média da sua vizinhança menos o valor de uma constante [5].

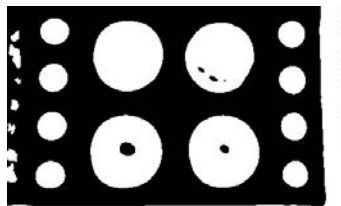


Figura 6.5: Resultado da binarização do sensor.

5. **Aplicar a *Hough Circle Transform*** - Esta operação resulta num conjunto de círculos [5]. Esse conjunto é de seguida filtrado, sendo removidos quaisquer círculos que se intersectem uns com os outros.



(a) Resultado da *Hough Circle Transform*.



(b) Resultado final.

Figura 6.6: Círculos detetados antes e depois da filtragem.

6. **Validação da Detecção** - A validação do processo da deteção depende do número de círculos restantes e se estes se encontram completamente dentro da zona esperada (parte do sensor mostrada na figura 6.6).

Caso os doze círculos pretendidos sejam corretamente detetados, é mostrada ao utilizador no ecrã a imagem do sensor com as zonas assinaladas, como podemos observar na

figura 6.7. Isto se este tiver selecionado o modo de depuração nas definições da aplicação. Quando o modo de depuração não está ativo, é mostrado ao utilizador diretamente ao resultado da análise após a deteção. Caso a deteção não tenha sucesso, é capturada outra imagem do sensor para que se possa repetir o processo.



Figura 6.7: Ecrã da aplicação no momento da deteção do sensor. Quando a aplicação se encontra no modo de depuração as regiões detetadas estão assinaladas a verde.

## 6.2 Calibração da Cor

O algoritmo selecionado para a calibração das cores da imagem do sensor, já descrito na secção 4.2.1, é baseado em oito zonas de referência de diferentes cores. Para incluir este algoritmo na aplicação foi apenas necessário adaptar a deteção do sensor e o cálculo das matrizes de correção já existentes para acomodar as quatro referências adicionais.

Para validar a consistência dos resultados obtidos na fase de seleção do algoritmo, foram realizados testes através da aplicação, utilizando os biossensores finais de diversas concentrações. Os sensores foram capturados sob as mesmas três condições de iluminação referidas e com os mesmos dois dispositivos. Podemos observar um exemplo nas figuras 6.8 e 6.9, sendo que os resultados integrais estão disponíveis na secção B.3.

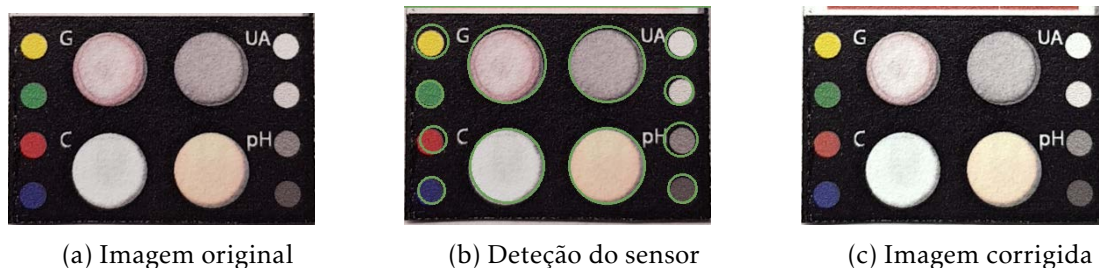


Figura 6.8: Detecção e correção do sensor através da aplicação com o *smartphone* Samsung S8 sob luz Led

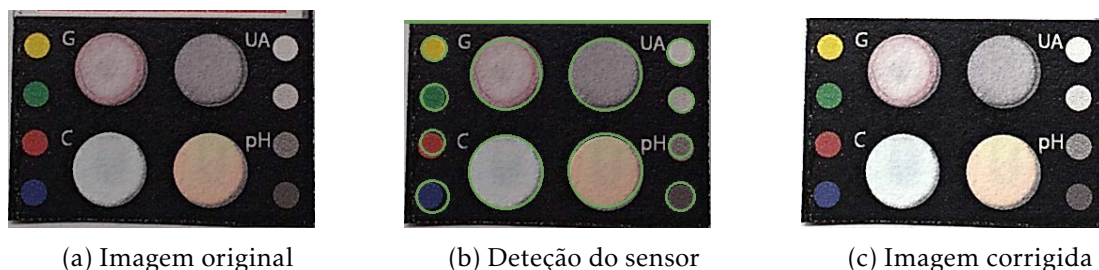


Figura 6.9: Detecção e correção do sensor através da aplicação com o *smartphone* Motorola G3 sob luz Led

Através da observação das imagens obtidas é possível verificar a consistência da correção não só em relação às imagens obtidas durante a seleção do algoritmo, mas também ao comparar as imagens de diferentes iluminações e dispositivos móveis.

No trabalho anterior [37] foram observados problemas com a correção dos sensores vazios, nomeadamente a existência de um tom azul ou verde na imagem após a correção. Estas tonalidades inesperadas levaram à obtenção de resultados incorretos na análise. Como é possível observar na figura 6.10, a nova correção implementada não resulta em sensores com este tipo de tonalidades, pelo que este problema aparenta ter sido solucionado com a adição das novas cores de referência.

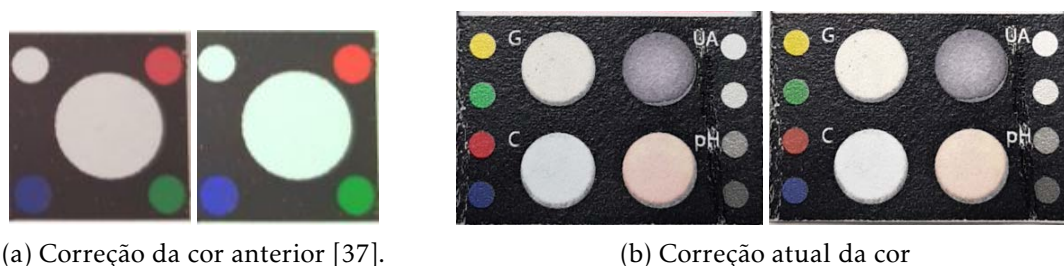


Figura 6.10: Comparação entre a correção implementada anteriormente e a correção atual, para o sensor vazio sob luz fluorescente.

## 6.3 Análise Colorimétrica

### 6.3.1 Implementação

Todo o processo de estudo e seleção dos melhores classificadores para a análise colorimétrica (secção 4.1) foi feito com recurso à biblioteca *Scikit-Learn* para a linguagem de programação *Python*. Para permitir a implementação dos modelos na aplicação, todo o processamento de dados necessário e o treino dos três classificadores selecionados, foi reproduzido na linguagem C++ através de funções da biblioteca *OpenCV*.

Os classificadores selecionados com base nos resultados detalhados na secção 5.1 são:

- **Classificação da glucose em quatro níveis de referência:**
  - **Em jejum** - Algoritmo KNN baseado na média, desvio padrão e assimetria dos três componentes do modelo YUV, sem aplicação de PCA.
  - **Duas horas após a refeição** - Algoritmo KNN baseado na média dos componentes L e S do modelo HLS, sem aplicação de PCA (figura 6.11).
- **Classificação do ácido úrico em três níveis de referência** - Algoritmo SVM baseado na média dos componentes H e S dos modelos de cor HSV e HLS, com aplicação de PCA (figuras 6.12 e 6.13).

As representações gráficas com recurso a pca e as regiões de decisão correspondentes todos os modelos testados estão disponíveis nos anexos G e F, respetivamente.

Um dos principais objetivos da aplicação é a rápida análise da cor da amostra e apresentação do resultado. Assim, não seria eficiente realizar o processamento dos dados e o treino dos classificadores de cada vez que fosse necessário analisar uma nova amostra. Deste modo, os classificadores treinados e os resultados do PCA nos vários conjuntos de dados, são exportados para ficheiros incluídos na própria aplicação. Tal permite a sua reutilização para que em cada análise só seja processada a nova amostra que é depois aplicada diretamente ao classificador para previsão.

Em relação à classificação do ácido úrico, devido à inexistência de exemplos referentes ao sensor vazio, o classificador não está preparado para identificar amostras vazias. Para evitar a incerteza de resultados para esta situação em particular, foi estabelecido um valor de cor base para o estado vazio. Todas as amostras com valores RGB acima desta cor são consideradas vazias. Esta cor base foi estimada a partir da média da cor de amostras do sensor vazio sob diversas iluminações (solar, led, fluorescente). No entanto, esta não é a melhor solução pois a cor do sensor vazio é algo semelhante à apresentada para algumas concentrações. Deste modo é proposto no futuro a inclusão de exemplos vazios para que possa ser feita a classificação para estes casos.

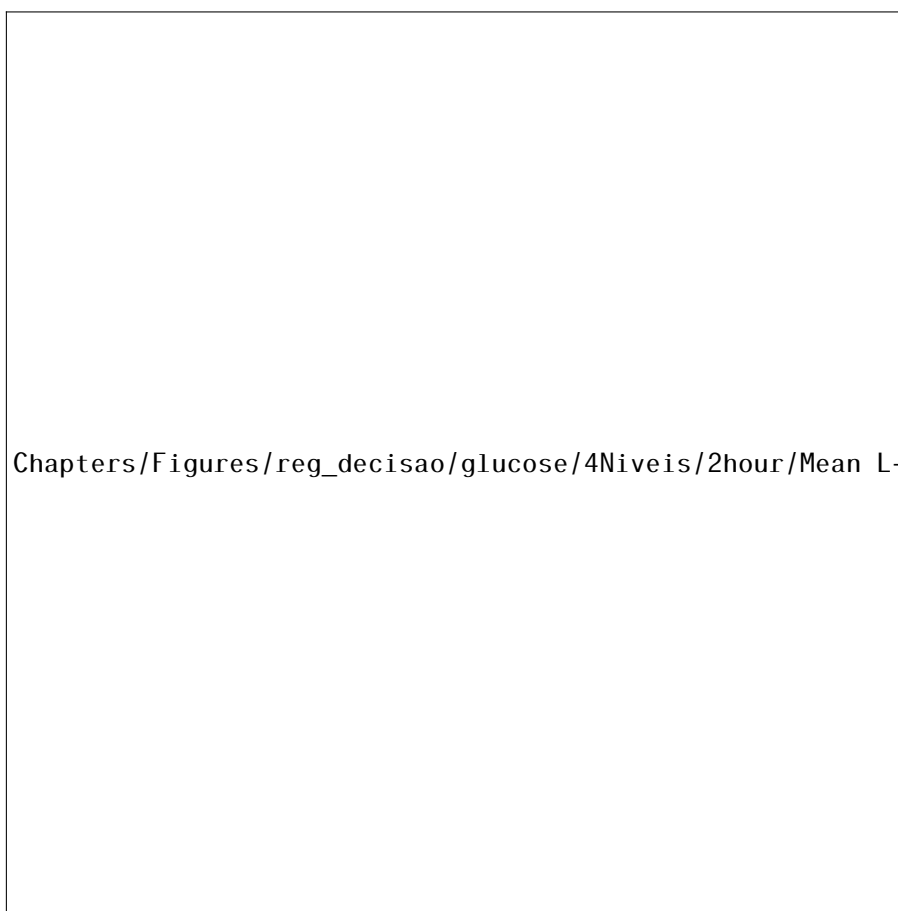


Figura 6.11: Região de decisão do algoritmo KNN para os quatro níveis da glucose, (duas horas após a refeição), baseado nas componentes L e S dos espaços de cor HLS (sem PCA).

### 6.3.2 Procedimento da Análise

Para realizar uma análise, após a deteção do sensor é pedido ao utilizador que escolha entre estado de jejum ou duas horas após a refeição. Consoante a escolha são seleccionados o classificador e o modelo aplicado à glucose. A análise do ácido úrico apenas inclui um modelo, pelo que não é necessário qualquer seleção.

As imagens das zonas extraídas dos dois poços superiores do sensor são depois corrigidas de acordo o processo calibração seleccionado. Após a correção são calculados os descritores de cor necessários para a análise. Esses dados são depois normalizados e processados com o PCA, caso necessário, e introduzidos no classificador para obter a previsão do resultado. Por último, o resultado da classificação é apresentado no ecrã da aplicação, acompanhado da descrição da cor extraída de cada região (Figura 6.14).

### 6.3.3 Testes

As bibliotecas de aprendizagem automática utilizadas na seleção dos classificadores e na implementação são muito distintas. Assim, para verificar se o comportamento dos algoritmos se manteve como era esperado, optou-se por testar o sistema com algumas amostras

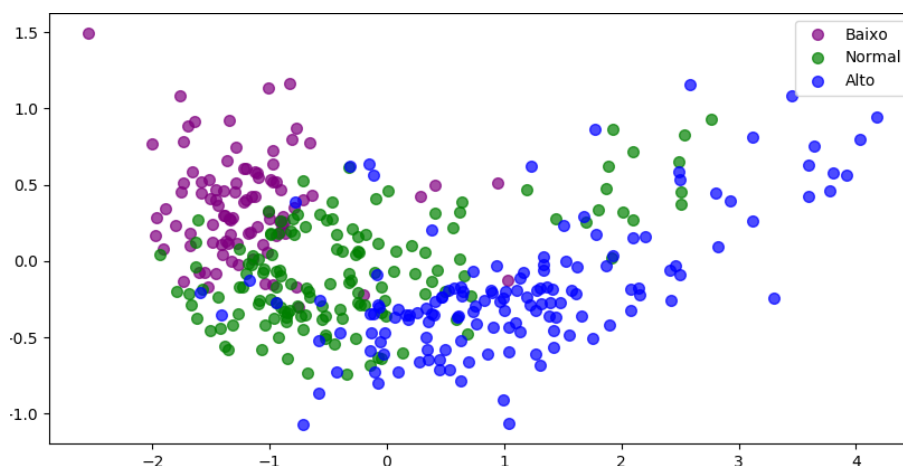


Figura 6.12: Visualização dos dados baseados nas H e S dos espaços de cor HLS e HSV, com recurso a PCA.

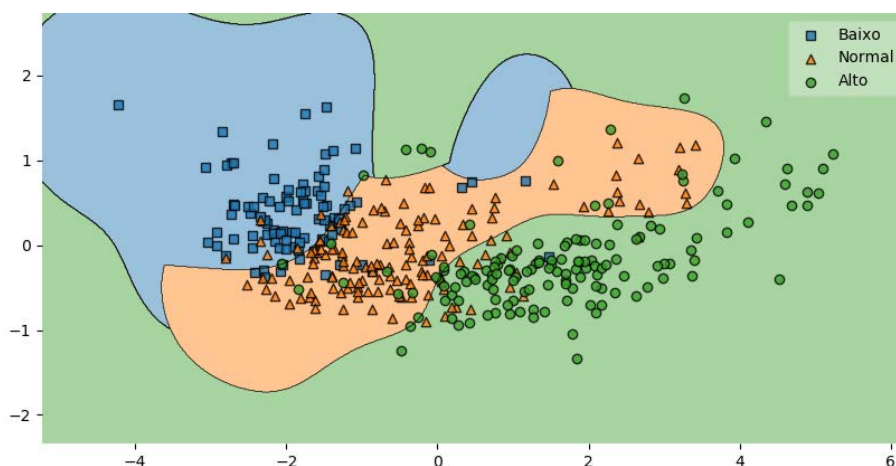


Figura 6.13: Região de decisão do algoritmo SVM para os três níveis de referência do ácido úrico, baseado nas componentes H e S dos espaços de cor HLS e HSV (com PCA).

iguais às usadas para o treino dos classificadores. Como era esperado, os classificadores identificaram todas as amostras corretamente.

Foram ainda feitos testes com aplicação sob os três tipos de iluminação diferentes (solar, fluorescente e led) com recurso a novos biossensores fornecidos pelo CENIMAT. Estes sensores consistem em amostras referentes não só a concentrações já conhecidas mas também a valores não usados durante o processo de treino:

- **Ácido úrico** - Vazio, 4 mM (nível alto), 1 mM (nível alto) e 1.5 mM (nível alto).
- **Glucose** - Vazio, 7.5 mM (jejum: diabetes, duas horas: normal/pré-diabetes), 1.5 mM (jejum: normal/pré-diabetes, duas horas: Hipoglicemia), 15 mM (jejum e duas horas: diabetes).

Independentemente do *smartphone* utilizado, a aplicação classificou corretamente todos os exemplos do ácido úrico. Em relação à glucose, apesar de conseguir classificar



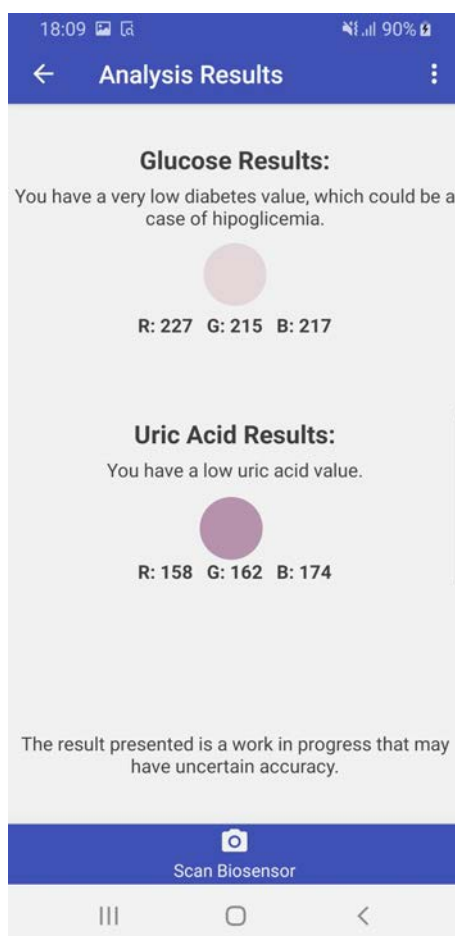


Figura 6.14: Ecrã da apresentação dos resultados da análise na aplicação

corretamente os exemplos referentes ao nível mais elevado (diabetes) e ao vazio, os classificadores não foram consistentes ao analisar sensores com as duas classes intermédias. Isto está de acordo com a maior capacidade do sistema para fazer uma análise qualitativa.

Contudo, estas amostras foram obtidas com recurso a processos laboratoriais diferentes dos usados nas amostras anteriores. Assim, as cores produzidas diferem um pouco das usadas no treino dos classificadores.

Outra condicionante está relacionada com o tempo decorrido desde a colocação do reagente até ao teste com o *smartphone*. Caso se demore mais do que alguns minutos a realizar o teste, a cor fica alterada e à medida que o tempo passa vai alterar cada vez mais. Esta situação dificulta a realização dos testes diretamente com a aplicação pois o laboratório onde são feitos os sensores apenas tem iluminação fluorescente e é necessário testar as outras iluminações. Deste modo, todos os testes foram feitos com algumas horas depois da introdução da amostra no sensor. Como se pode ver na figura 6.15, as cores são bastante diferentes, principalmente no caso do ácido úrico.

Assim, com exceção do sensor vazio, os resultados dos testes podem ter sido condicionados por estas situações.





(a) Algumas horas após a colocação da amostra.



(b) Logo após a colocação da amostra.

Figura 6.15: Comparação entre as cores das amostras logo após a colocação dos reagentes e algumas horas depois.



## CONCLUSÕES E TRABALHO FUTURO

De acordo com os objetivos da dissertação foi desenvolvida uma nova versão da aplicação capaz de analisar e classificar as duas substâncias presentes no novo biossensor (glucose e ácido úrico), através de algoritmos de aprendizagem automática.

Um dos objetivos consistia na adaptação do método de detecção da aplicação móvel ao novo biossensor. O sistema conseguiu detetar com sucesso todas as modificações de disposição e dimensões que foram sendo feitas ao sensor no decorrer da dissertação, incluindo a adição de zonas de referência para suportar o novo método de calibração implementado.

O método de calibração da cor do sensor proposto na versão anterior não era satisfatório. Para colmatar as lacunas da aplicação na calibração da cor, foi feito um estudo de vários algoritmos de calibração existentes. Foi ainda proposto um novo método de correção baseado na extensão do algoritmo anterior com a adição de cores de referência adicionais ao biossensor. Os resultados obtidos com este novo método foram muito superiores na redução das influências externas na cor, quando comparado com o implementado anteriormente.

Em relação à análise colorimétrica, para além de introduzir o ácido úrico na aplicação, pretendia-se repetir o processo desenvolvido no trabalho anterior, na tentativa de melhorar os resultados obtidos anteriormente na classificação da glucose através da introdução de um conjunto de dados maior. No entanto, o novo conjunto de dados fornecido para a glucose não era consistente em relação à cor correspondente a cada concentração. Assim, como não foi possível incluir todos os novos exemplos, apesar de maior, o conjunto tornou-se muito desequilibrado, prejudicando os resultados obtidos para a glucose.

Em comparação com o processo de análise anterior, procurou-se fornecer mais informação ao classificador a partir da cor da amostra, recorrendo não só à média da cor em diversos espaços de cor, mas também ao seu desvio padrão e assimetria. Foi então incluída

uma fase ao processo que procura estudar a cor das amostras, de modo a identificar as melhores *features* possíveis para cada modelo de dados.

Em relação à glucose, foram analisados os dois modelos testados anteriormente: as concentrações fornecidas e os quatro níveis de referência, bem como um terceiro modelo que faz a divisão total dos níveis de referência (5 níveis).

Apesar do aumento do conjunto de dados, infelizmente não foi possível obter melhores resultados para os modelos definidos no trabalho anterior. Os resultados para a classificação das concentrações mantiveram-se muito baixos. No entanto, foi possível obter resultados promissores (20% superiores) através de classificadores baseados nos três momentos da cor. Com o modelo dos quatro níveis, foram atingidos os resultados mais elevados para esta substância. Todavia, estes foram inferiores aos verificados anteriormente sendo que, no estado de duas horas após a refeição, foram cerca de 5% inferiores e no de jejum chegaram a ser 10% inferiores.

O novo modelo dos cinco níveis foi testado por fornecer uma informação mais detalhada ao utilizador. No entanto, apesar de os seus resultados serem favoráveis, foram ainda inferiores aos obtidos no modelo dos quatro níveis, pelo que não foi possível introduzi-lo na aplicação.

No que diz respeito à análise do ácido úrico, esta foi implementada de raiz na aplicação, seguindo os mesmos procedimentos da glucose. Durante o seu estudo procurou-se fazer uma classificação quantitativa das várias concentrações, mas tal como o observado na glucose, esta configuração obteve resultados muito insuficientes. Deste modo, foi criado um modelo a partir dos níveis de referência desta substância para o sexo masculino. Apesar deste não fornecer um resultado quantitativo, permite ao utilizador uma informação qualitativa relevante para o seu estado de saúde. Através deste modelo, foi possível obter resultados muito favoráveis para uma primeira análise.

Outros objetivos da dissertação procuravam que a introdução de novas substâncias fosse assegurada com alterações mínimas na interface da aplicação, mantendo assim a boa usabilidade conseguida nos testes de usabilidade realizados anteriormente. Assim, a introdução da nova substância na aplicação foi feita através de pequenas alterações ao ecrã onde eram já mostrados os resultados da análise da glucose.

Apesar de terem sido discutidas várias funcionalidades para melhorar a experiência do utilizador ao realizar a análise, foi dada prioridade à análise e classificação dos dados e ao desenvolvimento da calibração da cor. Estas funcionalidades são descritas como trabalho futuro, mas não foram introduzidas, pelo que não se considerou necessário realizar novos testes de usabilidade.

## 7.1 Trabalho Futuro

**Classificação da glucose:** No decorrer da conclusão desta dissertação foi desenvolvida no CENIMAT uma nova gama de cores para a classificação da glucose. Esta gama é mais semelhante à desenvolvida para o ácido úrico, visto que possui uma mudança de cor

distinguível a olho nu. Em comparação com a gama usada nesta dissertação (figura 7.1), esta aparenta ser mais promissora para a separação das várias concentrações. Apresenta uma variação significativa em certas features de alguns espaços de cor, principalmente na componente V do espaço de cor HSV, quando comparada com a variação obtida na gama anterior.

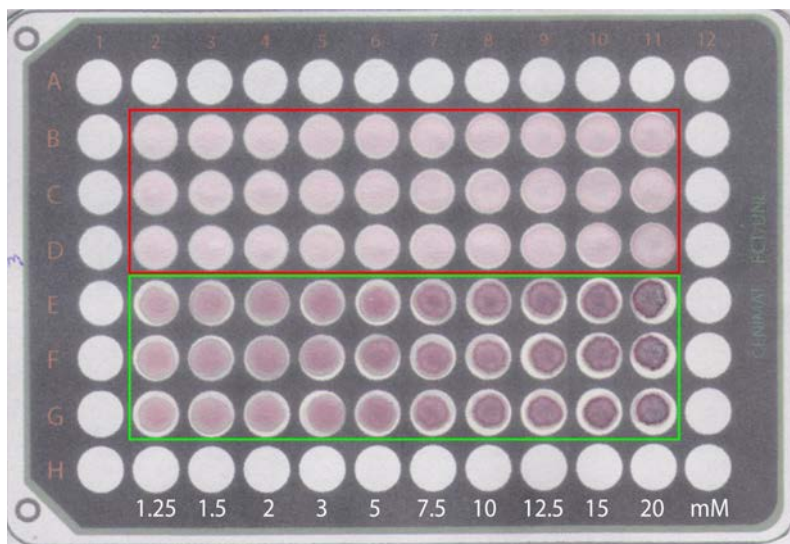


Figura 7.1: Comparação entre as gamas de cores disponíveis para a glucose: a usada nesta dissertação (vermelho) e a nova (verde).

Para procurar obter melhores resultados na classificação da glucose, seria então importante repetir o processo de análise realizado, de modo a fazer um estudo e seleção das *features* e classificadores que permitam obter os melhores resultados com esta gama de cores. Outro fator seria a utilização de um conjunto de dados equilibrado, com um maior número de exemplos por concentração do que os fornecidos para esta dissertação.

Entretanto, caso a análise das concentrações não consiga atingir resultados suficientemente elevados para permitir a análise quantitativa, o ideal seria obter valores que permitissem introduzir na aplicação o modelo dos cinco níveis de referência, visto que este fornece uma informação mais útil e clara ao utilizador do que o existente.

**Classificação do ácido úrico:** A primeira análise do ácido úrico mostrou resultados promissores na classificação dos níveis de referência. No entanto, para procurar melhorar os resultados para os níveis ou até para a classificação quantitativa das concentrações, seria importante aumentar o tamanho do conjunto de dados disponível, mantendo o equilíbrio já existente entre as várias classes.

Uma lacuna existente na análise desta substância é o facto de o estado vazio do sensor não estar incluído na classificação, o que a torna algo limitada. Tal ocorre, pois, não foram fornecidos exemplos para o estado em que o sensor está vazio (sem reagente). Assim, para evitar resultados inesperados ao testar um sensor vazio é então necessário

que seja fornecido um número de exemplos semelhante ao das outras concentrações, mas correspondentes ao estado vazio do sensor.

Caso os resultados obtidos com um maior conjunto de dados sejam mais favoráveis seria ainda importante implementar o modelo dos níveis de referência correspondentes ao sexo feminino, de modo a que a análise possa abranger todos os possíveis utilizadores.

**Análise colorimétrica:** Neste tipo de análise, como a única fonte de informação disponível é a cor da amostra, as gamas de cores disponibilizadas influenciam muito os resultados. Apesar de haver diferenças visíveis entre as amostras das concentrações mais elevadas e mais baixas, a variação da cor entre as concentrações intermédias é reduzida, não sendo visível a olho nu. Este detalhe torna a separação deste tipo de dados mais complicada.

Até agora, devido à necessidade da implementação, todos os algoritmos estudados no contexto desta aplicação foram algoritmos de classificação comuns, presentes na biblioteca *OpenCV*. Deste modo, pode ser interessante efetuar um estudo mais aprofundado no âmbito da aprendizagem automática, de forma a perceber se há alguma alternativa que permita atingir melhores resultados na classificação deste tipo de dados.

**Aplicação Móvel:** Para que a aplicação tenha mais utilidade para o utilizador que efetua a análise, seria importante adicionar funcionalidades que tornem todo o processo o mais simples e claro para o utilizador, procurando obter o melhor resultado possível para a análise.

Seria então interessante modificar a aplicação para que acompanhe o utilizador durante todo o processo da análise, desde a colocação do reagente no biossensor até à deteção e apresentação do resultado. A aplicação deve mostrar ao utilizador onde colocar a amostra no sensor. De seguida, um temporizador indicaria o tempo ótimo para obter a melhor cor possível para a análise. Com esse tempo esgotado, seria indicado ao utilizador que pode então iniciar a deteção do sensor para obter o resultado.

Ao fornecer esta informação adicional ao utilizador é possível reduzir a influência do indivíduo no resultado da análise.

## BIBLIOGRAFIA

- [1] E. Alpaydm. *Introduction to machine learning*. en. 2. ed. Adaptive computation and machine learning. Cambridge, Mass.: MIT Press, 2010. ISBN: 978-0-262-01243-0.
- [2] A. Azeroual e K. Afdel. “Fast Image Edge Detection based on Faber Schauder Wavelet and Otsu Threshold”. en. Em: *Heliyon* 3.12 (dez. de 2017). DOI: 10.1016/j.heliyon.2017.e00485.
- [3] B. Berg, B. Cortazar, D. Tseng, H. Ozkan, S. Feng, Q. Wei, R. Y.-L. Chan, J. Burbano, Q. Farooqui, M. Lewinski, D. D. Carlo, O. B. Garner e A. Ozcan. “Cellphone-Based Hand-Held Microplate Reader for Point-of-Care Testing of Enzyme-Linked Immunosorbent Assays”. Em: *ACS Nano* 9.8 (jul. de 2015), pp. 7857–7866. DOI: 10.1021/acsnano.5b03203.
- [4] G. Bradski. “The OpenCV Library”. Em: *Dr. Dobb’s Journal of Software Tools* (2000). URL: <https://opencv.org/>.
- [5] G. R. Bradski e A. Kaehler. *Learning OpenCV: computer vision with the OpenCV library*. eng. 1. ed. Sebastopol, CA: O’Reilly Media, 2008. ISBN: 978-0-596-51613-0.
- [6] D. Cheng, B. Price, S. Cohen e M. S. Brown. “Beyond White: Ground Truth Colors for Color Constancy Correction”. en. Em: *2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile: IEEE, dez. de 2015, pp. 298–306. DOI: 10.1109/ICCV.2015.42.
- [7] A. F. Coskun, J. Wong, D. Khodadadi, R. Nagi, A. Tey e A. Ozcan. “A personalized food allergen testing platform on a cellphone”. Em: *Lab Chip* 13.4 (2013), pp. 636–640. DOI: 10.1039/c2lc41152k.
- [8] J. Ferrão. “Aplicação da tecnologia de baixo custo Lab-on-Paper a biossensores colorimétricos não enzimáticos para a detecção de glucose”. Tese de mestrado. Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2016. URL: <http://hdl.handle.net/10362/26414>.
- [9] Q. Fu, Z. Wu, F. Xu, X. Li, C. Yao, M. Xu, L. Sheng, S. Yu e Y. Tang. “A portable smart phone-based plasmonic nanosensor readout platform that measures transmitted light intensities of nanosubstrates using an ambient light sensor”. Em: *Lab on a Chip* 16.10 (2016), pp. 1927–1933. DOI: 10.1039/c6lc00083e.

- [10] M. Habekost. "Which color differencing equation should be used?" en. Em: *International Circular of Graphic Education and Research* 6 (2013), pp. 20–33.
- [11] C. wei Hsu, C. chung Chang e C. jen Lin. "A practical guide to support vector classification". Em: (2016). URL: <https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf> (acedido em 09/07/2019).
- [12] J. Hu, S. Wang, L. Wang, F. Li, B. Pingguan-Murphy, T. J. Lu e F. Xu. "Advances in paper-based point-of-care diagnostics". Em: *Biosensors and Bioelectronics* 54 (abr. de 2014), pp. 585–597. DOI: 10.1016/j.bios.2013.10.075.
- [13] J. F. Hughes, ed. *Computer graphics: principles and practice*. eng. 3. ed. Upper Saddle River, NJ: Addison-Wesley, 2013. ISBN: 978-0-321-39952-6.
- [14] M.-Y. Jia, Q.-S. Wu, H. Li, Y. Zhang, Y.-F. Guan e L. Feng. "The calibration of cellphone camera-based colorimetric sensor array and its application in the determination of glucose in urine". Em: *Biosensors and Bioelectronics* 74 (dez. de 2015), pp. 1029–1037. DOI: 10.1016/j.bios.2015.07.072.
- [15] H. C. Karaimer e M. S. Brown. "Improving Color Reproduction Accuracy on Cameras". Em: Salt Lake City, UT: IEEE, jun. de 2018, pp. 6440–6449. DOI: 10.1109/CVPR.2018.00674.
- [16] H. Kim, O. Awofeso, S. Choi, Y. Jung e E. Bae. "Colorimetric analysis of saliva–alcohol test strips by smartphone-based instruments using machine-learning algorithms". en. Em: *Applied Optics* 56.1 (jan. de 2017), p. 84. DOI: 10.1364/AO.56.000084.
- [17] A. Koh, D. Kang, Y. Xue, S. Lee, R. M. Pielak, J. Kim, T. Hwang, S. Min, A. Banks, P. Bastien, M. C. Manco, L. Wang, K. R. Ammann, K.-I. Jang, P. Won, S. Han, R. Ghaffari, U. Paik, M. J. Slepian, G. Balooch, Y. Huang e J. A. Rogers. "A soft, wearable microfluidic device for the capture, storage, and colorimetric sensing of sweat". Em: *Science Translational Medicine* 8.366 (nov. de 2016), 366ra165–366ra165. DOI: 10.1126/scitranslmed.aaf2593.
- [18] R. M. Kumar e K. Sreekumar. "A survey on image feature descriptors". Em: *International Journal of Computer Science and Information Technologies* 5.6 (2014), pp. 7668–7673. URL: <http://ijcsit.com/docs/Volume%205/vol5issue06/ijcsit20140506168.pdf> (acedido em 24/02/2019).
- [19] W. Lee e W. Woo. "Real-time Color Correction for Marker-based Augmented Reality Applications". en. Em: *Proceedings of the Third International Workshop on Ubiquitous Virtual Reality (IWUVR2009)*. Mawson Lakes, S.A.: University of South Australia, 2009, pp. 32–35. ISBN: 978-0-9806098-0-6.
- [20] N. Lopez-Ruiz, V. F. Curto, M. M. Erenas, F. Benito-Lopez, D. Diamond, A. J. Palma e L. F. Capitan-Vallvey. "Smartphone-Based Simultaneous pH and Nitrite Colorimetric Determination for Paper Microfluidic Devices". Em: *Analytical Chemistry* 86.19 (set. de 2014), pp. 9554–9562. DOI: 10.1021/ac5019205.



- [21] C.-M. Ma, W.-S. Yang e B.-W. C. . “How the Parameters of K-nearest Neighbor Algorithm Impact on the Best Classification Accuracy: In Case of Parkinson Dataset”. Em: *Journal of Applied Sciences* 14.2 (fev. de 2014), pp. 171–176. DOI: 10.3923/jas.2014.171.176.
- [22] A. W. Martinez, S. T. Phillips, G. M. Whitesides e E. Carrilho. “Diagnostics for the Developing World: Microfluidic Paper-Based Analytical Devices”. Em: *Analytical Chemistry* 82.1 (jan. de 2010), pp. 3–10. DOI: 10.1021/ac9013989.
- [23] C. J. Musto, S. H. Lim e K. S. Suslick. “Colorimetric Detection and Identification of Natural and Artificial Sweeteners”. en. Em: *Analytical Chemistry* 81.15 (ago. de 2009), pp. 6526–6533. DOI: 10.1021/ac901019g.
- [24] A. Y. Mutlu, V. Kılıç, G. K. Özdemir, A. Bayram, N. Horzum e M. E. Solmaz. “Smartphone-based colorimetric detection via machine learning”. Em: *The Analyst* 142.13 (2017), pp. 2434–2441. DOI: 10.1039/c7an00741h.
- [25] V. Oncescu, D. O'Dell e D. Erickson. “Smartphone based health accessory for colorimetric detection of biomarkers in sweat and saliva”. Em: *Lab on a Chip* 13.16 (2013), p. 3232. DOI: 10.1039/c3lc50431j.
- [26] *OpenCV: Canny Edge Detector*. URL: [https://docs.opencv.org/master/da/d5c/tutorial\\_canny\\_detector.html](https://docs.opencv.org/master/da/d5c/tutorial_canny_detector.html) (acedido em 18/02/2019).
- [27] *OpenCV: Color conversions*. URL: [https://docs.opencv.org/4.1.0/de/d25/imgproc\\_color\\_conversions.html](https://docs.opencv.org/4.1.0/de/d25/imgproc_color_conversions.html) (acedido em 20/09/2019).
- [28] W. H. Organization. *Global status report on noncommunicable diseases 2010*. zh. World Health Organization, 2011. ISBN: 978-92-4-456422-6.
- [29] W. H. Organization. *Increasing access to diagnostics through technology transfer and local production*. en. World Health Organization, 2011. ISBN: 978-92-4-150237-5.
- [30] W. H. Organization. *Noncommunicable diseases country profiles 2018*. en. World Health Organization, 2018. ISBN: 978-92-4-151462-0.
- [31] J. M. Park e Y. L. Murphey. “Edge Detection in Grayscale, Color, and Range Images”. en. Em: *Wiley Encyclopedia of Computer Science and Engineering*. Ed. por B. W. Wah. Hoboken, NJ, USA: John Wiley & Sons, Inc., abr. de 2008. ISBN: 978-0-470-05011-8. DOI: 10.1002/9780470050118.ecse603. (Acedido em 05/12/2019).
- [32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot e d. Duchesnay. “Scikit-learn: Machine Learning in Python”. Em: *Journal of Machine Learning Research* 12.Oct (2011), pp. 2825–2830.
- [33] T. Pinheiro. “Multiplex, Enzyme-free, Colorimetric Paper-based Device for the Measurement of Glucose, Uric Acid and Cholesterol”. Tese de mestrado. Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2018. URL: <http://hdl.handle.net/10362/56405>.

- [34] A. A. P. dos Diabéticos de Portugal. *ABC da Diabetes*. 2019. URL: <https://apdp.pt/diabetes/abc-da-diabetes/> (acedido em 16/08/2019).
- [35] C. Poynton. "Frequently asked questions about color". en. Em: (), p. 24. URL: <http://poynton.ca/PDFs/ColorFAQ.pdf> (acedido em 24/02/2019).
- [36] A. Roda, M. Guardigli, D. Calabria, M. M. Calabretta, L. Cevenini e E. Michelini. "A 3D-printed device for a smartphone-based chemiluminescence biosensor for lactate in oral fluid and sweat". Em: *The Analyst* 139.24 (set. de 2014), pp. 6494–6501. DOI: 10.1039/c4an01612b.
- [37] A. Rodrigues. "Aplicação Móvel de Processamento de Imagem para Análise Quantitativa de Biossensores". Tese de mestrado. Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2018. URL: <http://hdl.handle.net/10362/61551>.
- [38] A. Rodrigues, N. Correia e E. Fortunato. "Mellitus: A Smartphone Application for Image Processing and Colorimetric Analysis". Em: *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*. MUM 2018. Cairo, Egypt: ACM, 2018, pp. 449–455. ISBN: 978-1-4503-6594-9. DOI: 10.1145/3282894.3289738.
- [39] K. E. A. van de Sande, T Gevers e C. G. M. Snoek. "Evaluating Color Descriptors for Object and Scene Recognition". Em: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.9 (set. de 2010), pp. 1582–1596. DOI: 10.1109/tpami.2009.154.
- [40] G. Sharma, W. Wu e E. N. Dalal. "The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations". en. Em: *Color Research & Application* 30.1 (fev. de 2005), pp. 21–30. DOI: 10.1002/col.20070.
- [41] L. Shen, J. A. Hagen e I. Papautsky. "Point-of-care colorimetric detection with a smartphone". Em: *Lab on a Chip* 12.21 (2012), p. 4240. DOI: 10.1039/c2lc40741h.
- [42] E. Smith, C. Díaz-Torné, F. Perez-Ruiz e L. March. "Epidemiology of gout: An update". en. Em: *Best Practice & Research Clinical Rheumatology* 24.6 (dez. de 2010), pp. 811–827. ISSN: 15216942. DOI: 10.1016/j.berh.2010.10.004.
- [43] M. Sokolova e G. Lapalme. "A systematic analysis of performance measures for classification tasks". en. Em: *Information Processing & Management* 45.4 (jul. de 2009), pp. 427–437. ISSN: 03064573. DOI: 10.1016/j.ipm.2009.03.002.
- [44] W. S. Souza, M. A. S. de Oliveira, G. M. F. de Oliveira, D. P. de Santana e R. E. de Araujo. "Self-Referencing Method for Relative Color Intensity Analysis Using Mobile-Phone". Em: *Optics and Photonics Journal* 08.07 (2018), pp. 264–275. DOI: 10.4236/opj.2018.87022.
- [45] M. A. Stricker e M. Orengo. "Similarity of color images". Em: *Storage and Retrieval for Image and Video Databases III*. Vol. 2420. International Society for Optics e Photonics. SPIE, 1995, pp. 381–392. DOI: 10.1117/12.205308.

- 
- [46] J. Suykens e J. Vandewalle. “Least Squares Support Vector Machine Classifiers”. Em: *Neural Processing Letters* 9.3 (jun. de 1999), pp. 293–300. DOI: 10.1023/A:1018628609742.
- [47] R. Szeliski. *Computer Vision*. Texts in Computer Science. London: Springer London, 2011. ISBN: 978-1-84882-935-0. DOI: 10.1007/978-1-84882-935-0.
- [48] S. K. Vashist e J. H. T. Luong. “Smartphone-Based Point-of-Care Technologies for Mobile Healthcare”. en. Em: *Point-of-Care Technologies Enabling Next-Generation Healthcare Monitoring and Management*. Cham: Springer International Publishing, 2019, pp. 27–79. DOI: 10.1007/978-3-030-11416-9\_2.
- [49] S. K. Vashist, P. B. Lippa, L. Y. Yeo, A. Ozcan e J. H. Luong. “Emerging Technologies for Next-Generation Point-of-Care Testing”. en. Em: *Trends in Biotechnology* 33.11 (nov. de 2015), pp. 692–705. DOI: 10.1016/j.tibtech.2015.09.001.
- [50] E. Whitley e J. Ball. “Statistics review 1: Presenting and summarising data”. Em: *Critical Care* 6.1 (nov. de 2001), p. 66. DOI: 10.1186/cc1455.
- [51] *X-Rite: Colorimetric values for ColorChecker Family of Targets*. Jan. de 2010. URL: [https://xritephoto.com/ph\\_product\\_overview.aspx?ID=1257\&Action=Support\&SupportID=5159](https://xritephoto.com/ph_product_overview.aspx?ID=1257\&Action=Support\&SupportID=5159) (acedido em 13/08/2019).
- [52] A. K. Yetisen, J. Martinez-Hurtado, A. Garcia-Melendrez, F. da Cruz Vasconcellos e C. R. Lowe. “A smartphone algorithm with inter-phone repeatability for the analysis of colorimetric tests”. en. Em: *Sensors and Actuators B: Chemical* 196 (jun. de 2014), pp. 156–160. DOI: 10.1016/j.snb.2014.01.077.
- [53] A. K. Yetisen, M. S. Akram e C. R. Lowe. “Paper-based microfluidic point-of-care diagnostic devices”. Em: *Lab on a Chip* 13.12 (2013), p. 2210. DOI: 10.1039/c3lc50169h.
- [54] Y. Zhang, Y. Wu, Y. Zhang e A. Ozcan. “Color calibration and fusion of lens-free and mobile-phone microscopy images for high-resolution and accurate color reproduction”. Em: *Scientific Reports* 6.1 (jun. de 2016). DOI: 10.1038/srep27811.
- [55] F. Zhou, H. B.-L. Duh e M. Billinghurst. “Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR”. Em: *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 193–202. DOI: 10.1109/ISMAR.2008.4637362.





## RESULTADOS DA COMPARAÇÃO DA COR

Neste apêndice estão disponíveis os resultados dos testes realizados para a comparação e avaliação dos algoritmos de correção da cor do sensor, detalhados na seção 3.2.

As imagens do sensor, capturadas sob diferentes condições de iluminação (solar, led, fluorescente) e dispositivos (*Samsung S8*, *Motorola G3*), são comparadas entre si e com imagens de referência. A diferença entre as imagens é quantificada usando a métrica  $\Delta E_{2000}$  [40].

### A.1 Comparação da Correção das Imagens sob os Três Tipos de Iluminação Analisados

	Anterior	Modificado	Jia et al.	Lee et al.	Lopez-Ruiz et al.	Zhang et al.	Souza et al.
Luz Fluorescente	88.12	61.18	118.01	70.43	115.29	128.29	207.59
Luz Led	139.65	63.37	54.04	26.94	45.39	54.38	59.38
Luz Solar	38.46	32.76	71.70	52.05	31.38	191.98	178.95
Total	266.23	157.31	243.75	149.43	192.06	374.64	445.92

Tabela A.1: Soma do  $\Delta E_{2000}$  para a comparação entre as imagens obtidas com os dois modelos de *smartphone* analisados.

## A.2 Comparação da Correção das Imagens entre os Modelos de *Smartphone* Analisados

	Anterior	Modificado	Jia et al.	Lee et al.	Lopez-Ruiz et al.	Zhang et al.	Souza et al.
Motorola G3	365.01	275.77	314.48	218.13	255.28	611.73	294.48
Samsung S8	329.66	267.70	303.24	256.85	192.53	401.43	136.03
Total	694.67	543.47	617.72	474.98	447.82	1013.16	430.51

Tabela A.2: Soma do  $\Delta E_{2000}$  para a comparação entre as várias iluminações analisadas para cada modelo de *smartphone*.

## A.3 Comparação entre os Resultados da Correção e as Referências

	Anterior	Modificado	Jia et al.	Lee et al.	Lopez-Ruiz et al.	Zhang et al.	Souza et al.
Luz Fluorescente Motorola G3	415.15	90.66	109.24	133.23	147.23	165.19	63.90
Luz Fluorescente Samsung S8	377.01	101.67	93.46	161.58	174.39	119.97	20.60
Luz Led Motorola G3	584.57	266.53	201.77	273.97	267.17	105.27	16.66
Luz Led Samsung S8	492.09	116.43	90.28	155.76	148.67	94.49	11.80
Luz Solar Motorola G3	65.17	75.48	80.69	145.20	135.29	282.04	11.64
Luz Solar Samsung S8	368.58	69.11	91.00	160.35	137.41	141.26	12.40
Total Motorola G3	1064.88	432.67	391.70	552.39	549.69	552.50	92.21
Total Samsung S8	1237.68	287.21	274.74	477.69	460.47	355.72	44.80
Total	2302.56	719.89	666.44	1030.08	1010.16	908.23	137.01

Tabela A.3: Soma do  $\Delta E_{2000}$  para a comparação entre as imagens de referência e as imagens obtidas com os dois modelos de *smartphone* analisados.

### A.3. COMPARAÇÃO ENTRE OS RESULTADOS DA CORREÇÃO E AS REFERÊNCIAS

	Anterior	Modificado	Jia et al.	Lee et al.	Lopez-Ruiz et al.	Zhang et al.	Souza et al.
Luz Fluorescente Motorola G3	211.87	170.05	229.34	183.08	206.52	296.80	221.14
Luz Fluorescente Samsung S8	74.52	66.81	44.63	62.13	99.51	197.50	220.42
Luz Led Motorola G3	208.35	109.01	87.46	70.72	86.36	191.41	213.01
Luz Led Samsung S8	111.77	88.55	56.15	74.63	78.95	175.08	180.61
Luz Solar Motorola G3	109.01	86.08	78.63	73.21	97.43	263.26	184.34
Luz Solar Samsung S8	116.22	89.35	122.54	119.92	109.27	231.48	209.27
Total Motorola G3	529.23	365.13	395.43	327.01	390.31	751.46	618.49
Total Samsung S8	302.51	244.70	223.33	256.68	287.72	604.06	610.29
Total	831.74	609.83	618.76	583.69	678.03	1355.53	1228.78

Tabela A.4: Soma do  $\Delta E_{2000}$  para a comparação entre os sensores no formato digital e as imagens obtidas com os dois modelos de *smartphone* analisados.



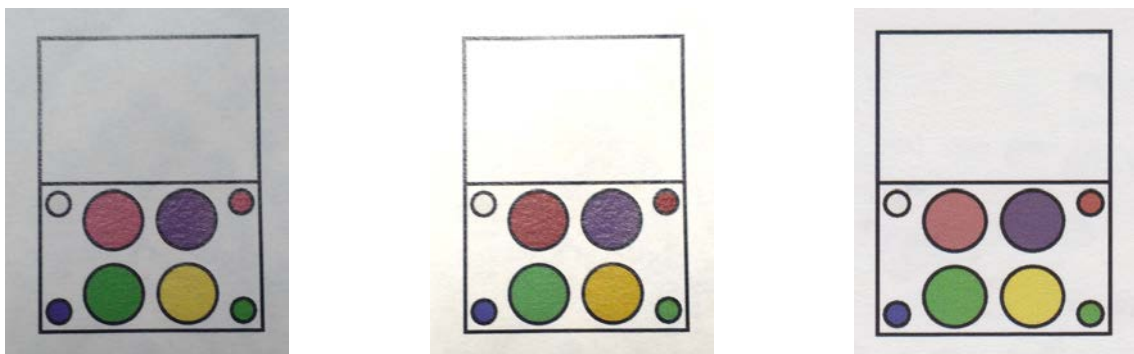


## RESULTADOS DA CORREÇÃO DA COR

Neste apêndice estão disponíveis os resultados da correção do biossensor com os vários algoritmos de calibração da cor, descritos na secção 3.2. Dois dos algoritmos testados não devolvem a correção integral da imagem, devolvendo apenas o valor da cor corrigida. Assim, para estes é apenas apresentado o *design* do sensor usado para realizar o teste. Por último, são apresentados os resultados obtidos nos testes usando a aplicação móvel, após a implementação do novo método de correção. Os sensores usados para realizar os diferentes testes foram capturados em condições de iluminação distintas (fluorescente, led e solar) com dois *smartphones* distintos: *Samsung S8* (Gama alta) e *Motorola G3* (Gama média).

### B.1 Resultados da Correção para Comparação dos Algoritmos de Calibração Estudados

#### B.1.1 Luz Fluorescente



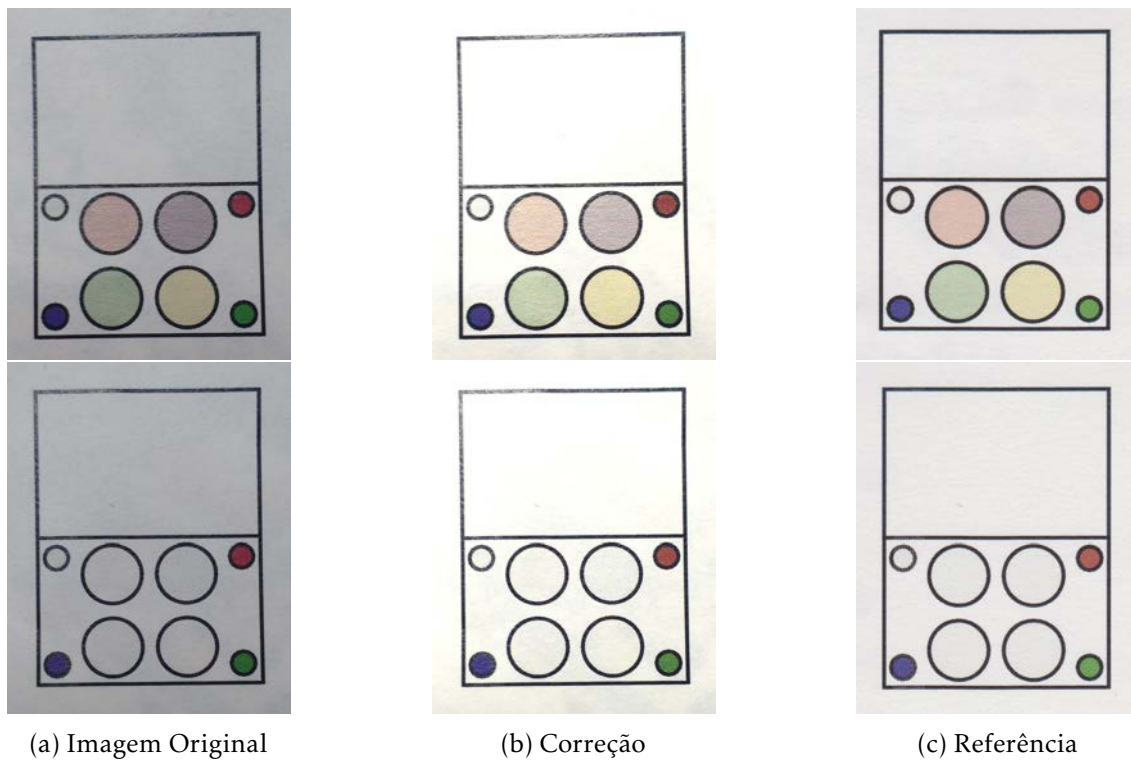
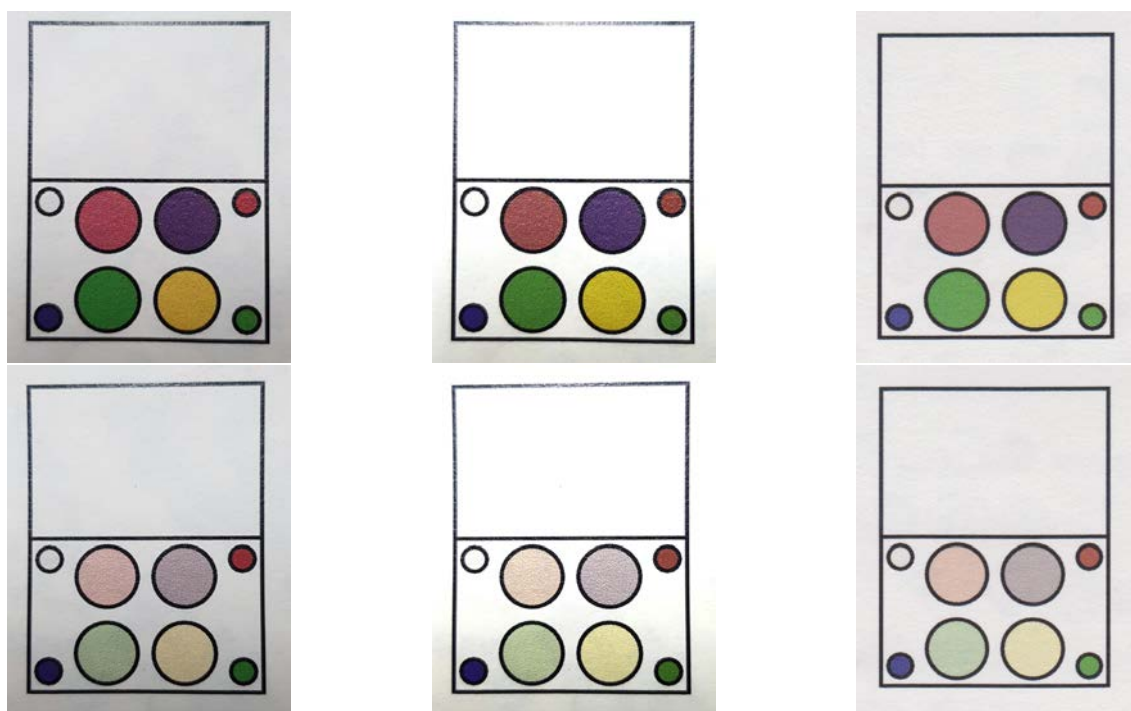


Figura B.1: Correção do sensor (cor intensa, cor média e vazia) fotografado pelo *smartphone Motorola G3* através do algoritmo da versão anterior da aplicação, proposto por Rodrigues [37], descrito na secção 3.2.3.



## B.1. RESULTADOS DA CORREÇÃO PARA COMPARAÇÃO DOS ALGORITMOS DE CALIBRAÇÃO ESTUDADOS

---



Figura B.1: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Samsung S8* através do algoritmo da versão anterior da aplicação, proposto por Rodrigues [37], descrito na secção 3.2.3.

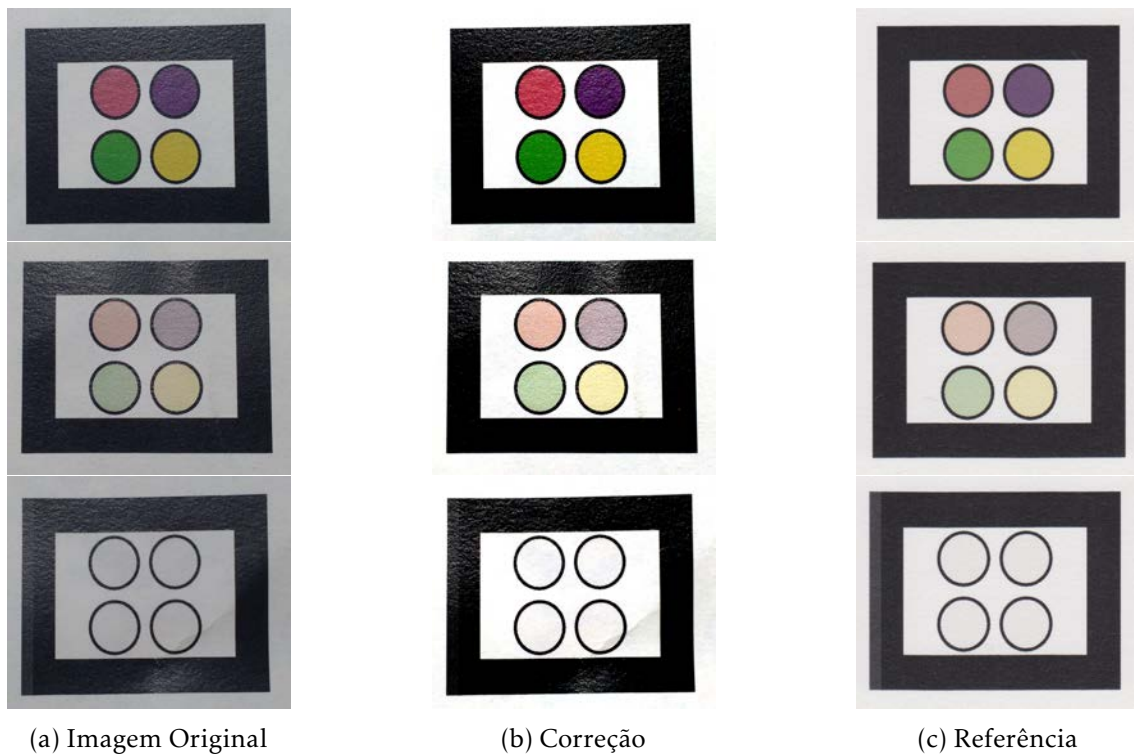


Figura B.2: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Motorola G3* através do algoritmo proposto por Lee e Woo [19], descrito na secção 3.2.2.2.

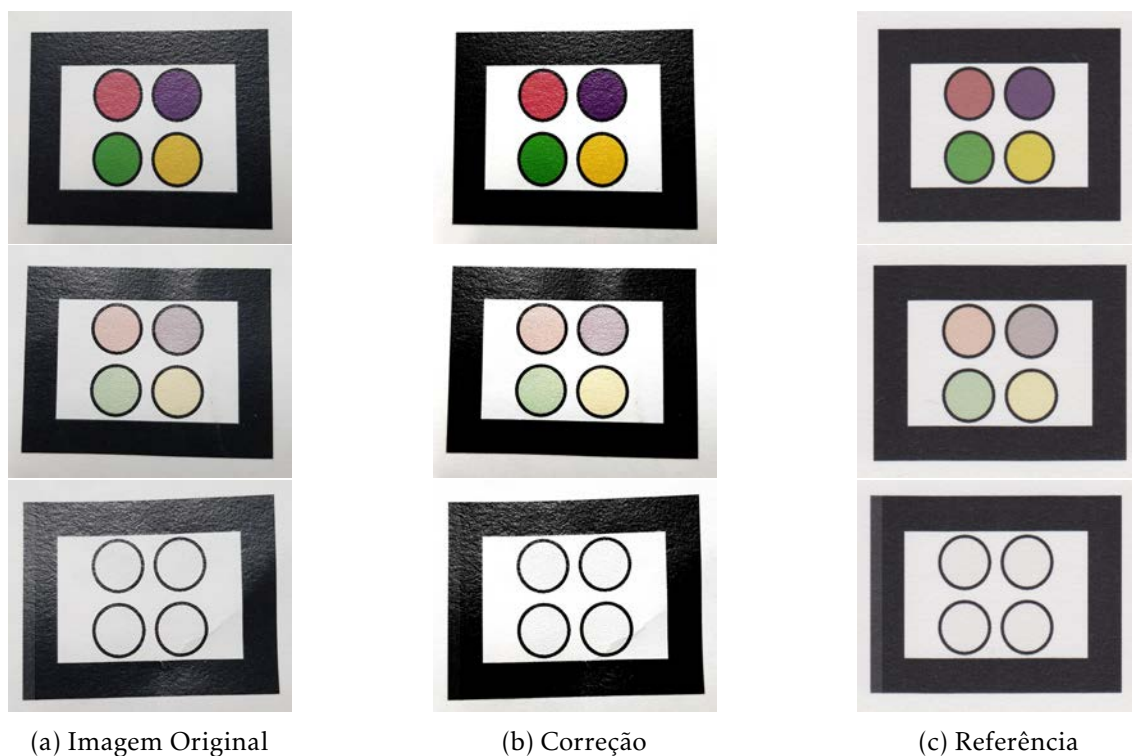
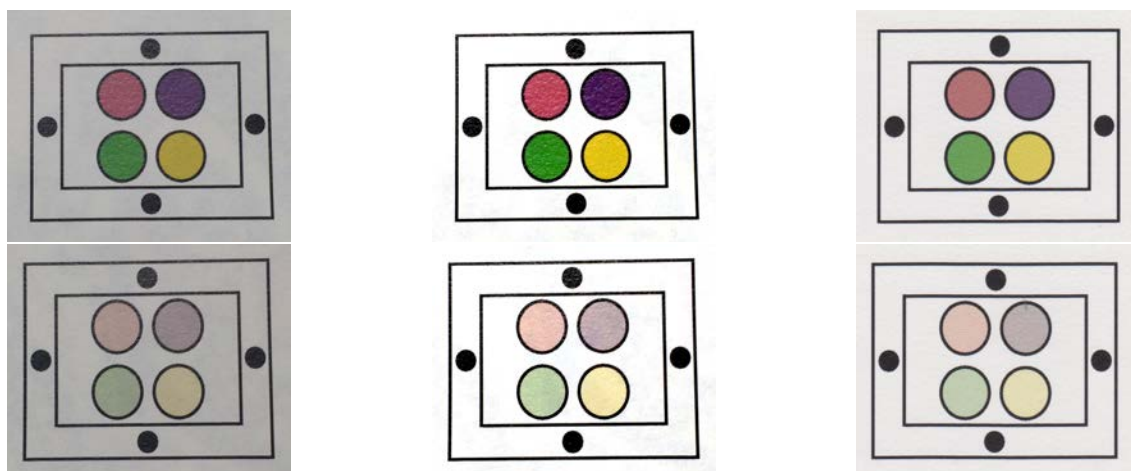


Figura B.3: Correção do sensor (cor intensa, cor média e vazia) fotografado pelo *smartphone Samsung S8* através do algoritmo proposto por Lee e Woo [19], descrito na secção 3.2.2.2.





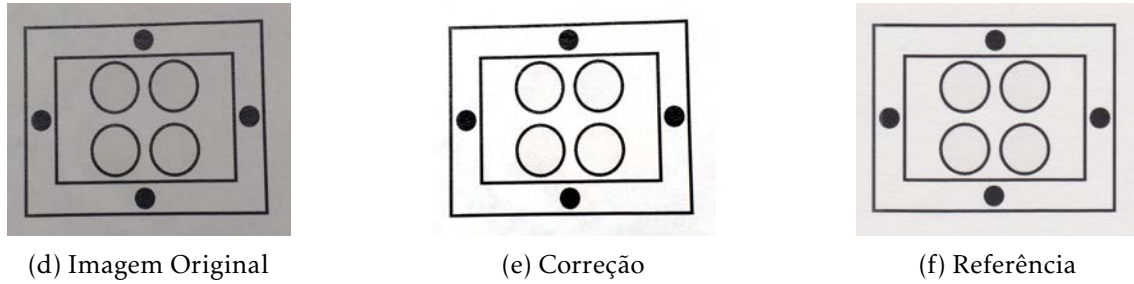


Figura B.3: Correção do sensor (cor intensa, cor média e vazia) fotografado pelo *smartphone Motorola G3* através do algoritmo proposto por Jia et al. [14], descrito em 3.2.2.1.

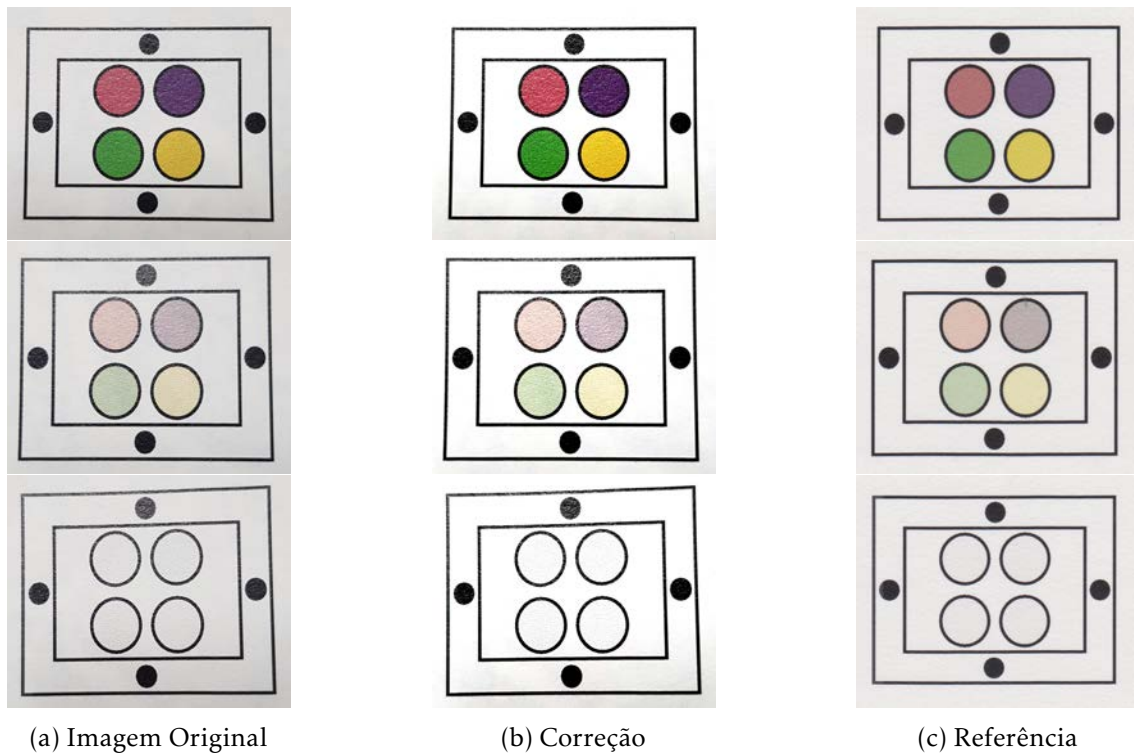


Figura B.4: Correção do sensor (cor intensa, cor média e vazia) fotografado pelo *smartphone Samsung S8* através do algoritmo proposto por Jia et al. [14], descrito em 3.2.2.1.



Figura B.5: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Motorola G3* através do algoritmo proposto por Zhang et al. [54], descrito em 3.2.2.4.



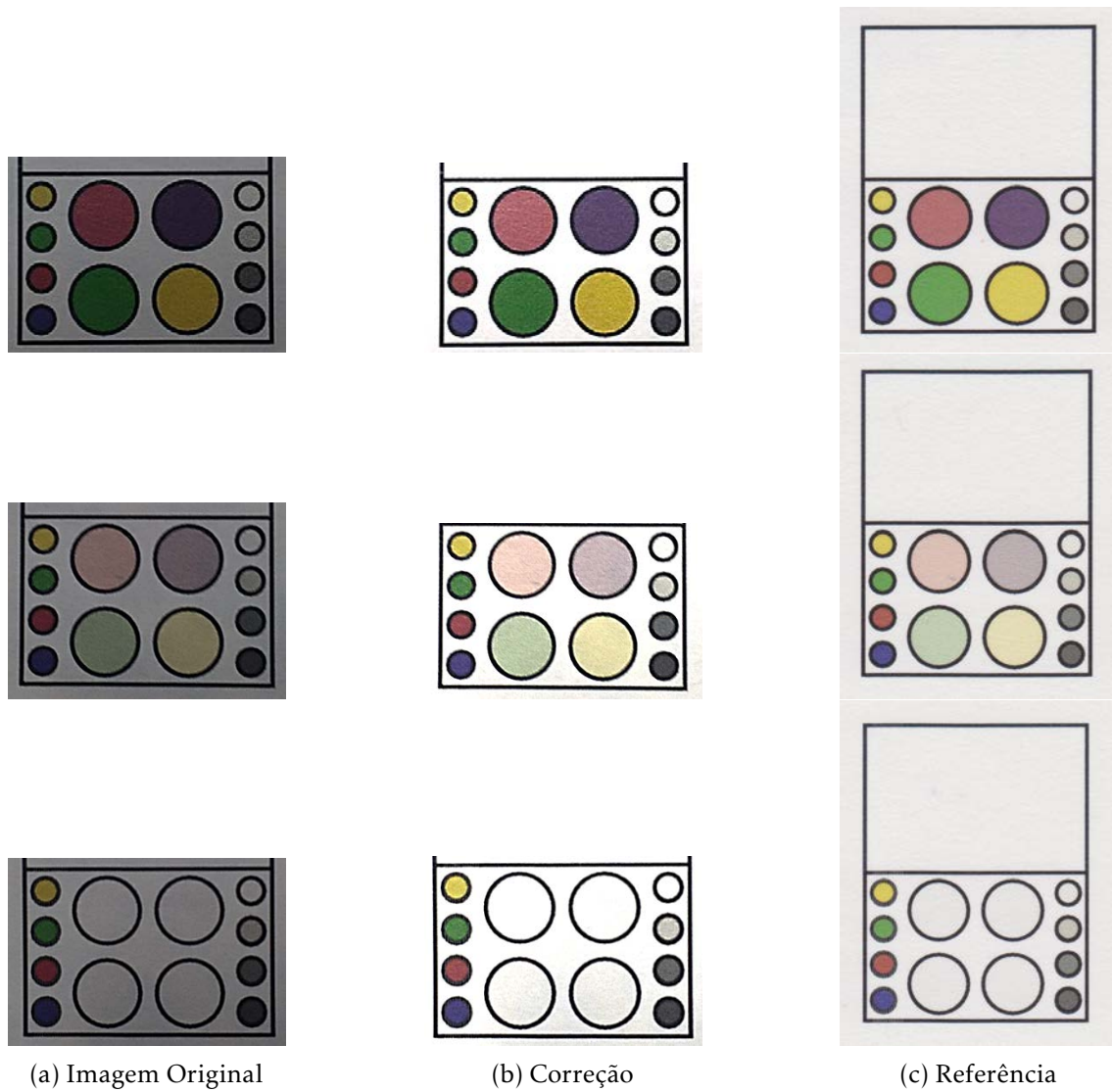


Figura B.7: Correção do sensor de fundo branco (cor intensa, cor média e vazio) fotografado pelo *smartphone Motorola G3* através do algoritmo modificado, descrito na secção 3.2.3.



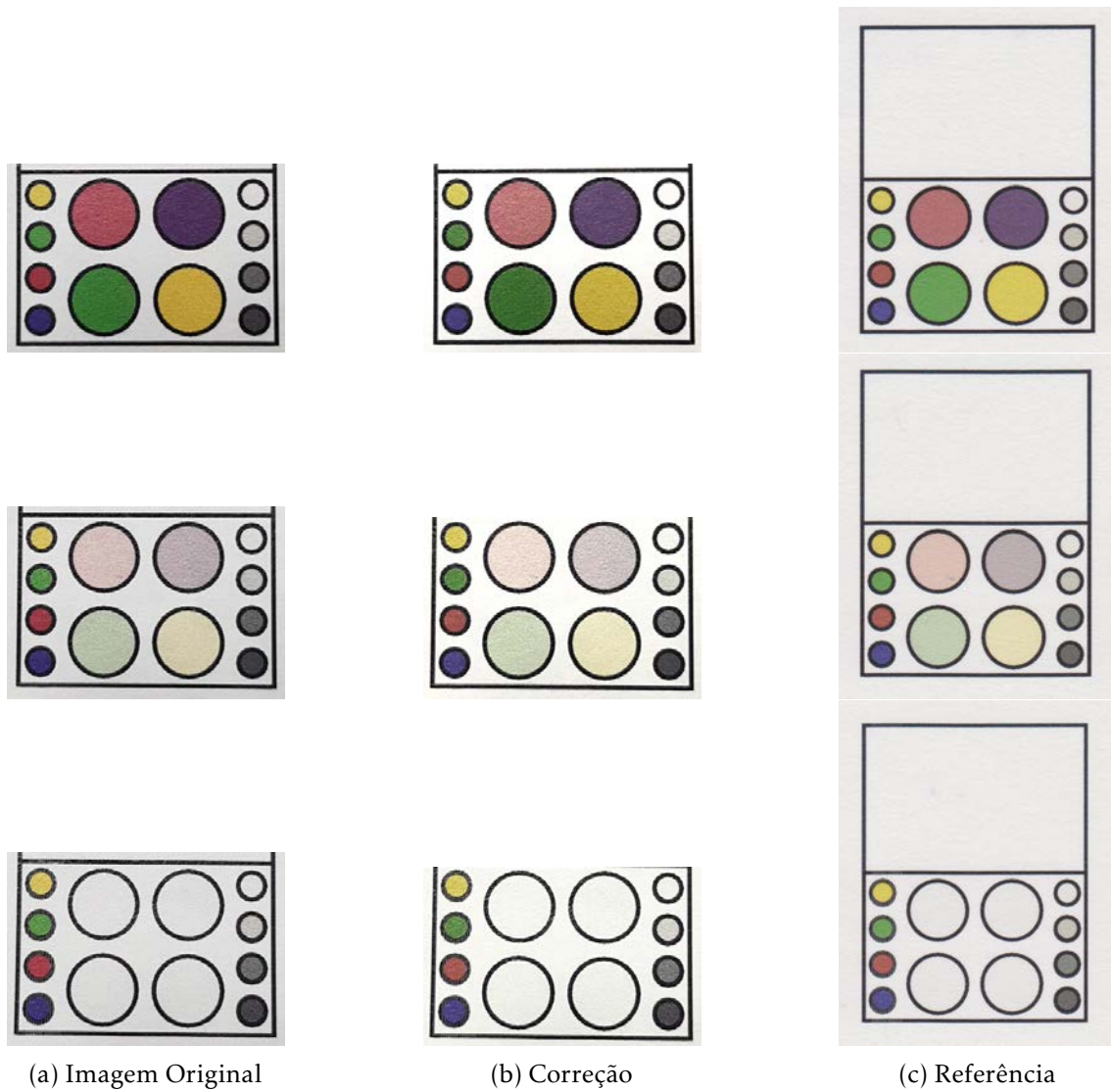


Figura B.8: Correção do sensor de fundo branco (cor intensa, cor média e vazio) fotografado pelo *smartphone Samsung S8* através do algoritmo modificado, descrito na secção 3.2.3.

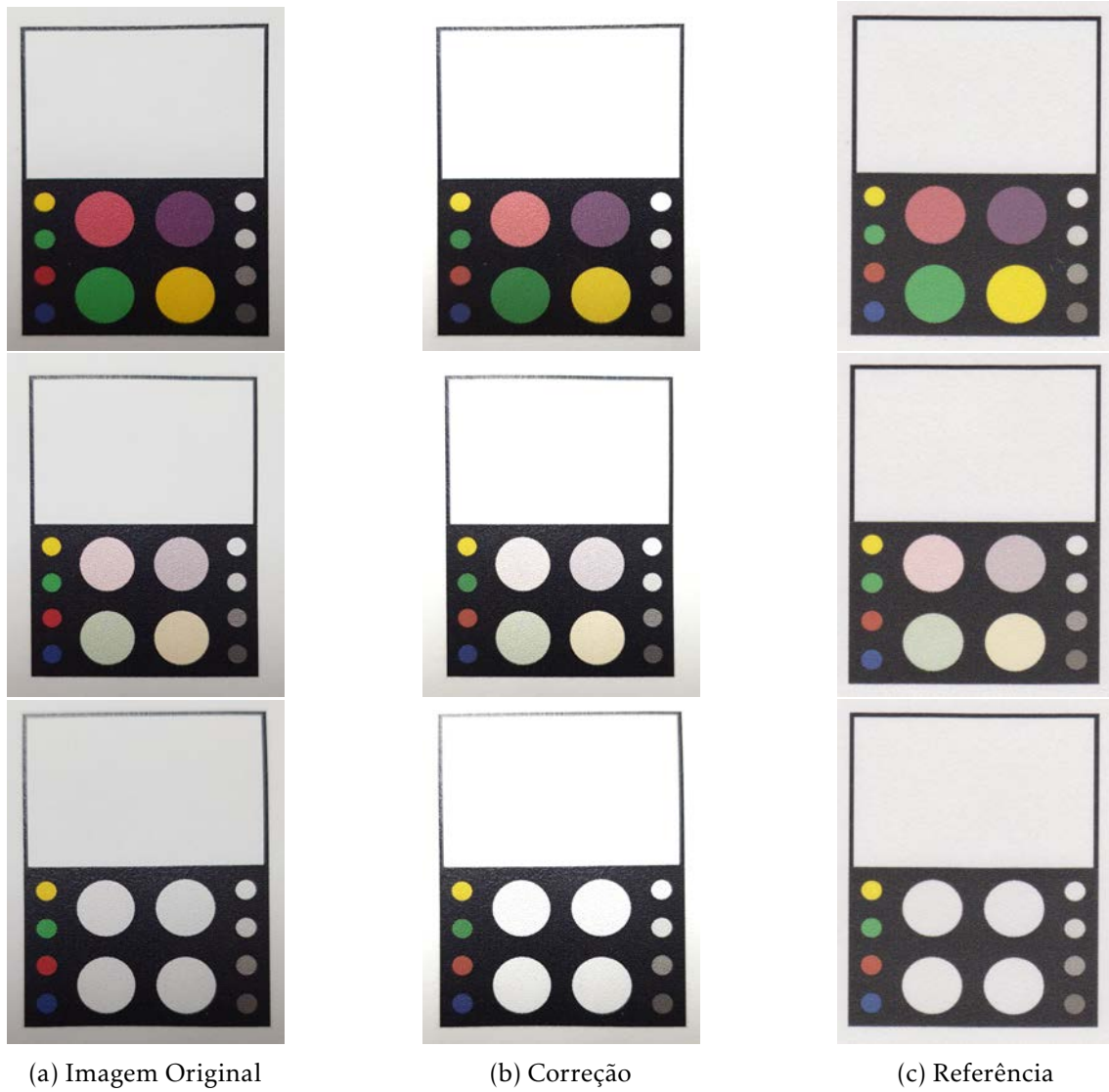


Figura B.9: Correção do sensor de fundo preto (cor intensa, cor média e vazio) fotografado pelo *smartphone Motorola G3* através do algoritmo modificado, descrito na secção 3.2.3.

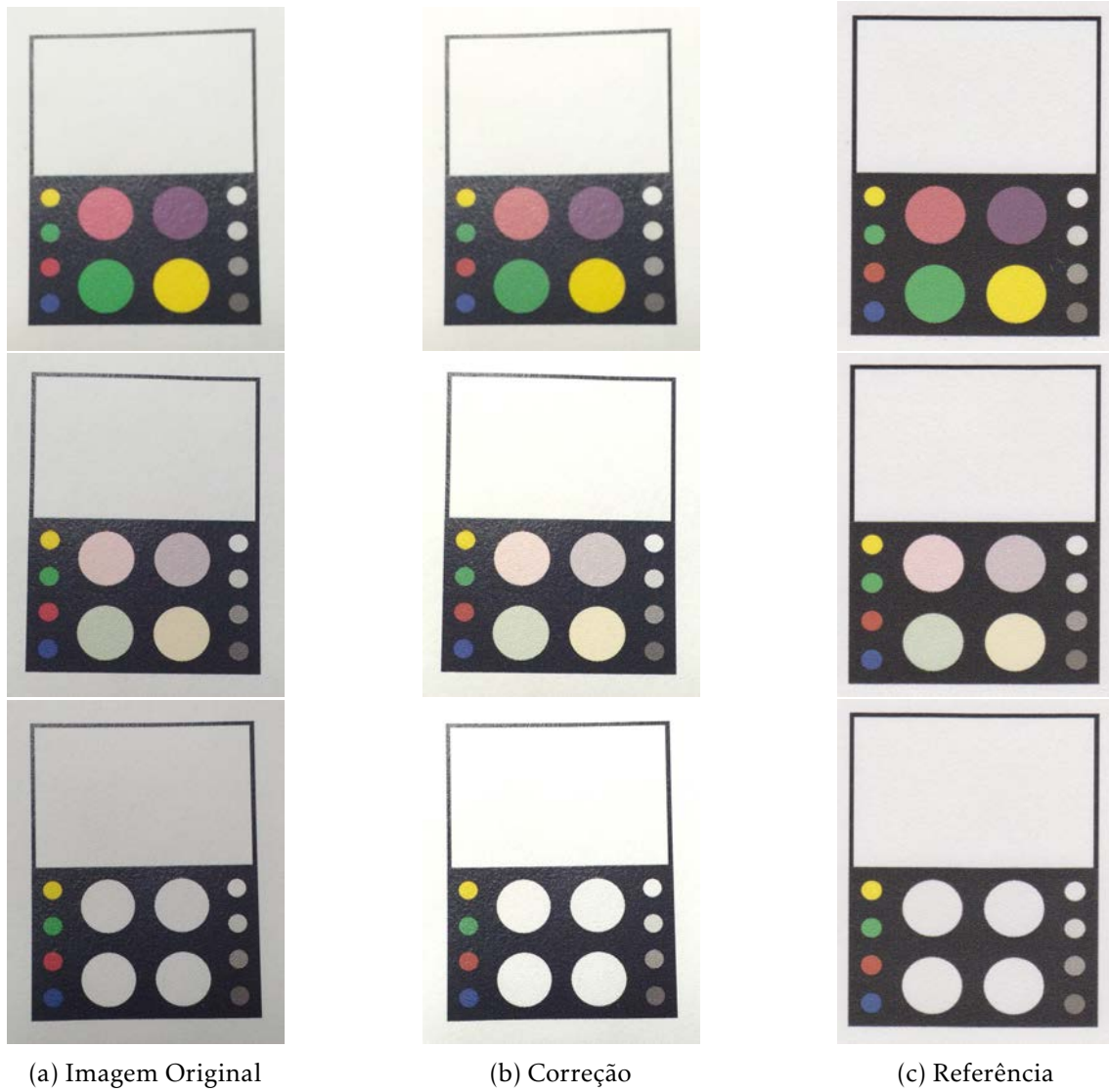


Figura B.10: Correção do sensor de fundo preto (cor intensa, cor média e vazio) fotografado pelo *smartphone Samsung S8* através do algoritmo modificado, descrito na secção 3.2.3.

### B.1.2 Luz Led

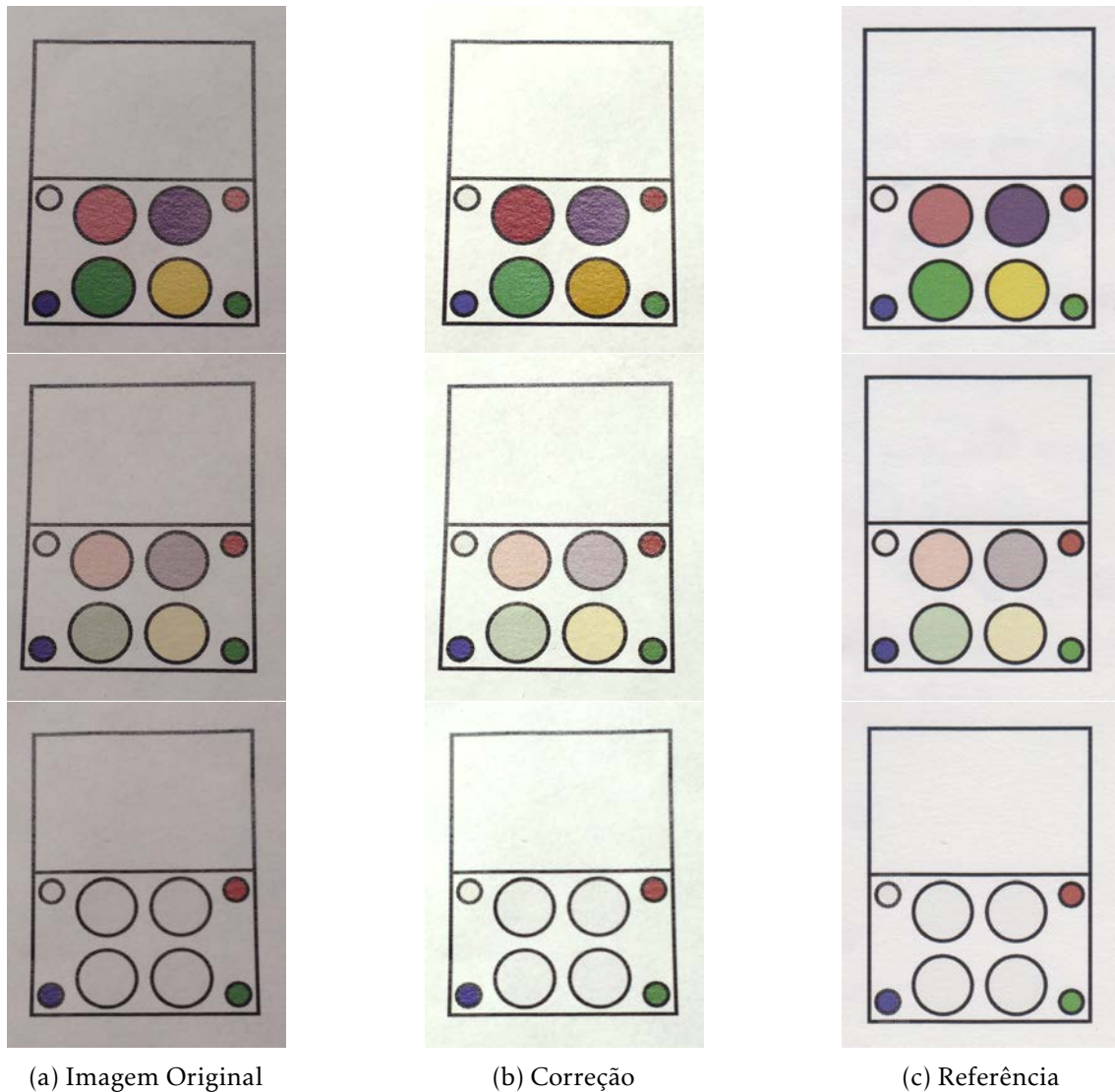


Figura B.11: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Motorola G3* através do algoritmo da versão anterior da aplicação, proposto por Rodrigues [37], descrito na secção 3.2.3.

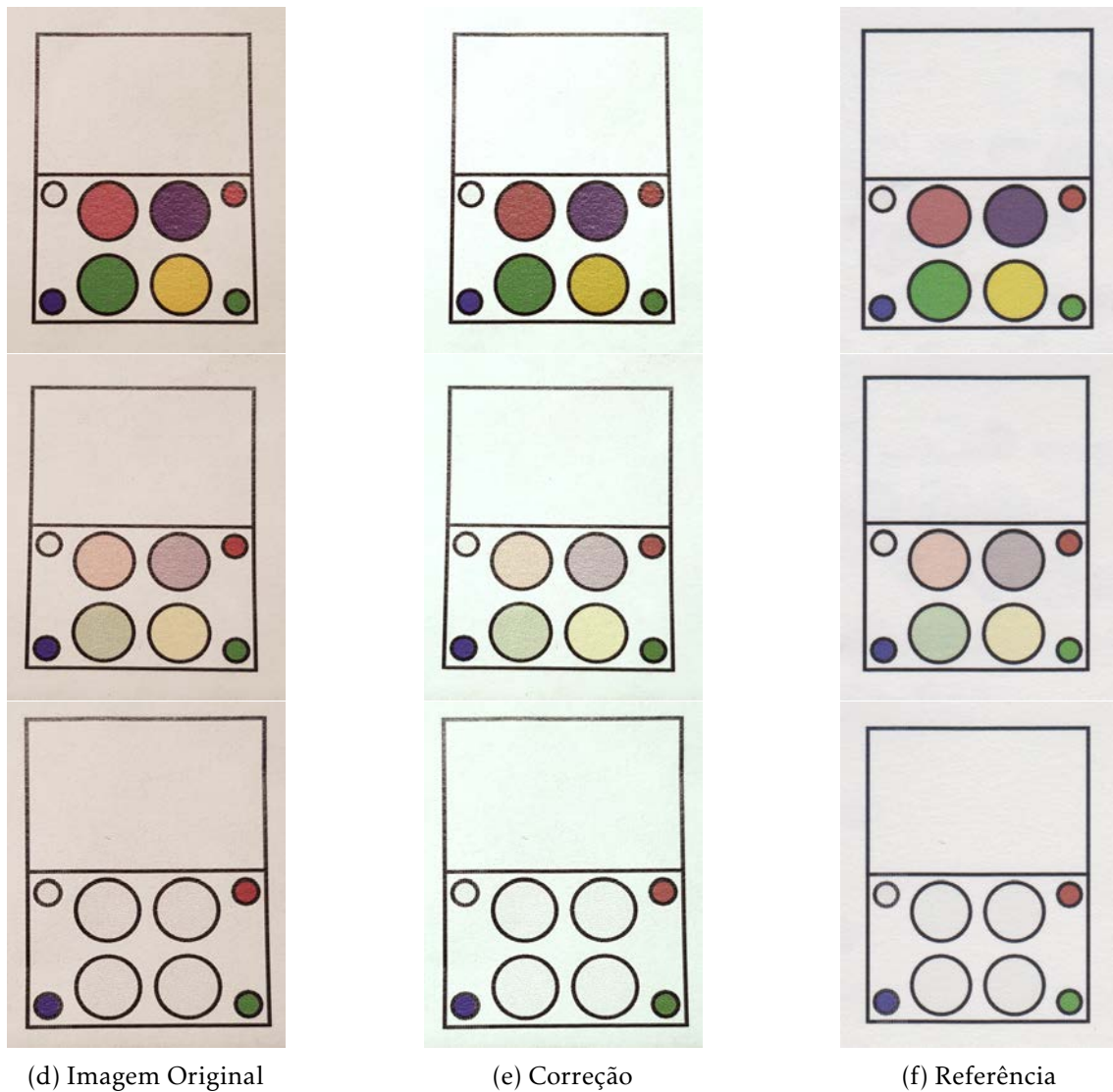


Figura B.11: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Samsung S8* através do algoritmo da versão anterior da aplicação, proposto por Rodrigues [37], descrito na secção 3.2.3.



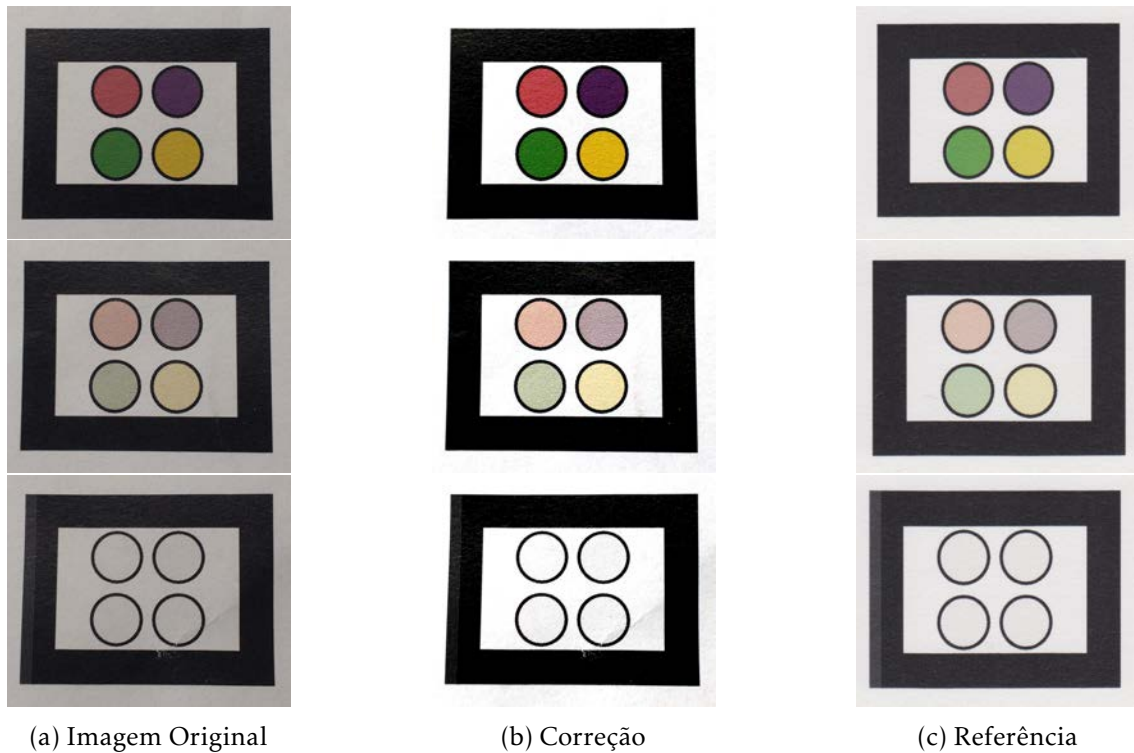


Figura B.12: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Motorola G3* através do algoritmo proposto por Lee e Woo [19], descrito na secção 3.2.2.2.



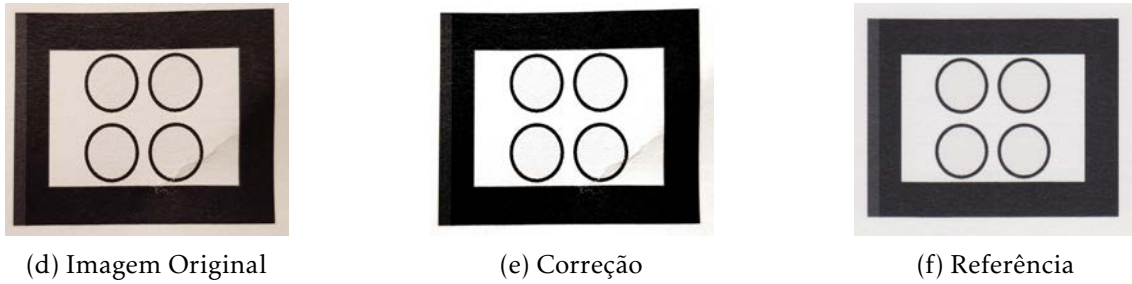


Figura B.12: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Samsung S8* através do algoritmo proposto por Lee e Woo [19], descrito na secção 3.2.2.2.

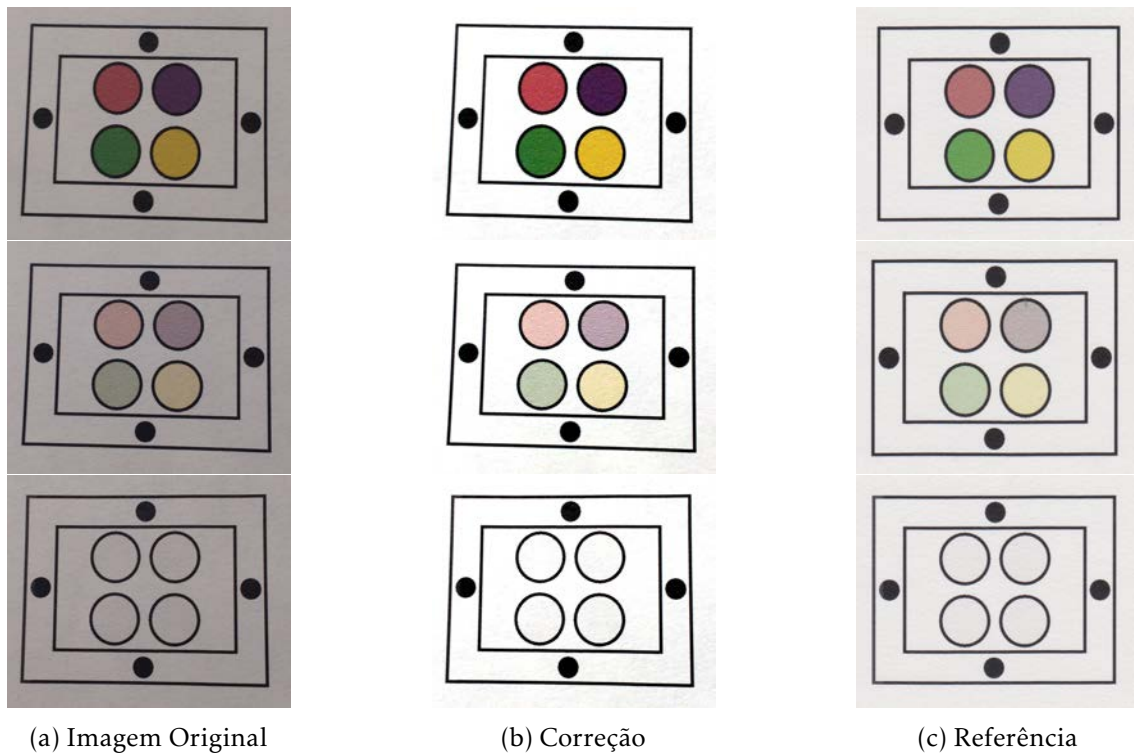


Figura B.13: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Motorola G3* através do algoritmo proposto por Jia et al. [14], descrito em 3.2.2.1.

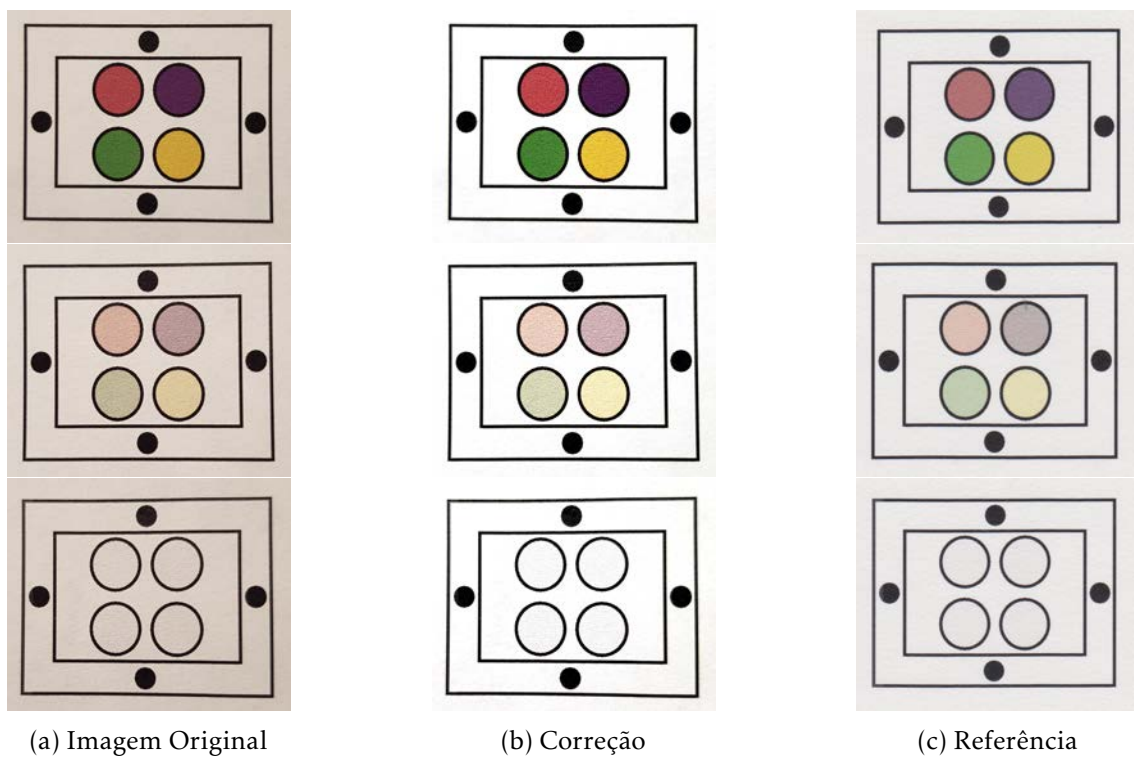


Figura B.14: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Samsung S8* através do algoritmo proposto por Jia et al. [14], descrito em 3.2.2.1.



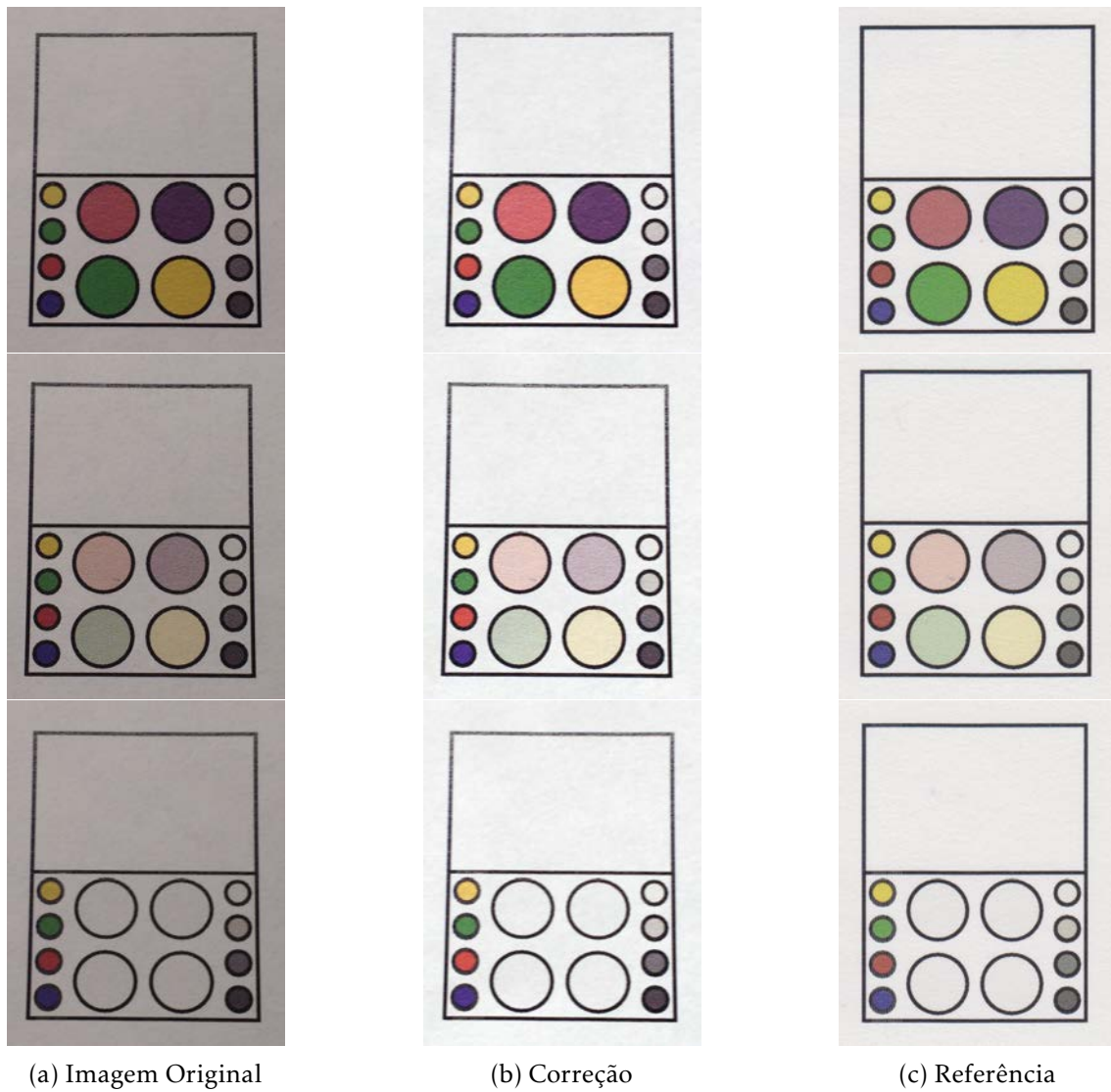


Figura B.15: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Motorola G3* através do algoritmo proposto por Zhang et al. [54], descrito em 3.2.2.4.



Figura B.16: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Samsung S8* através do algoritmo proposto por Zhang et al. [54], descrito em 3.2.2.4.

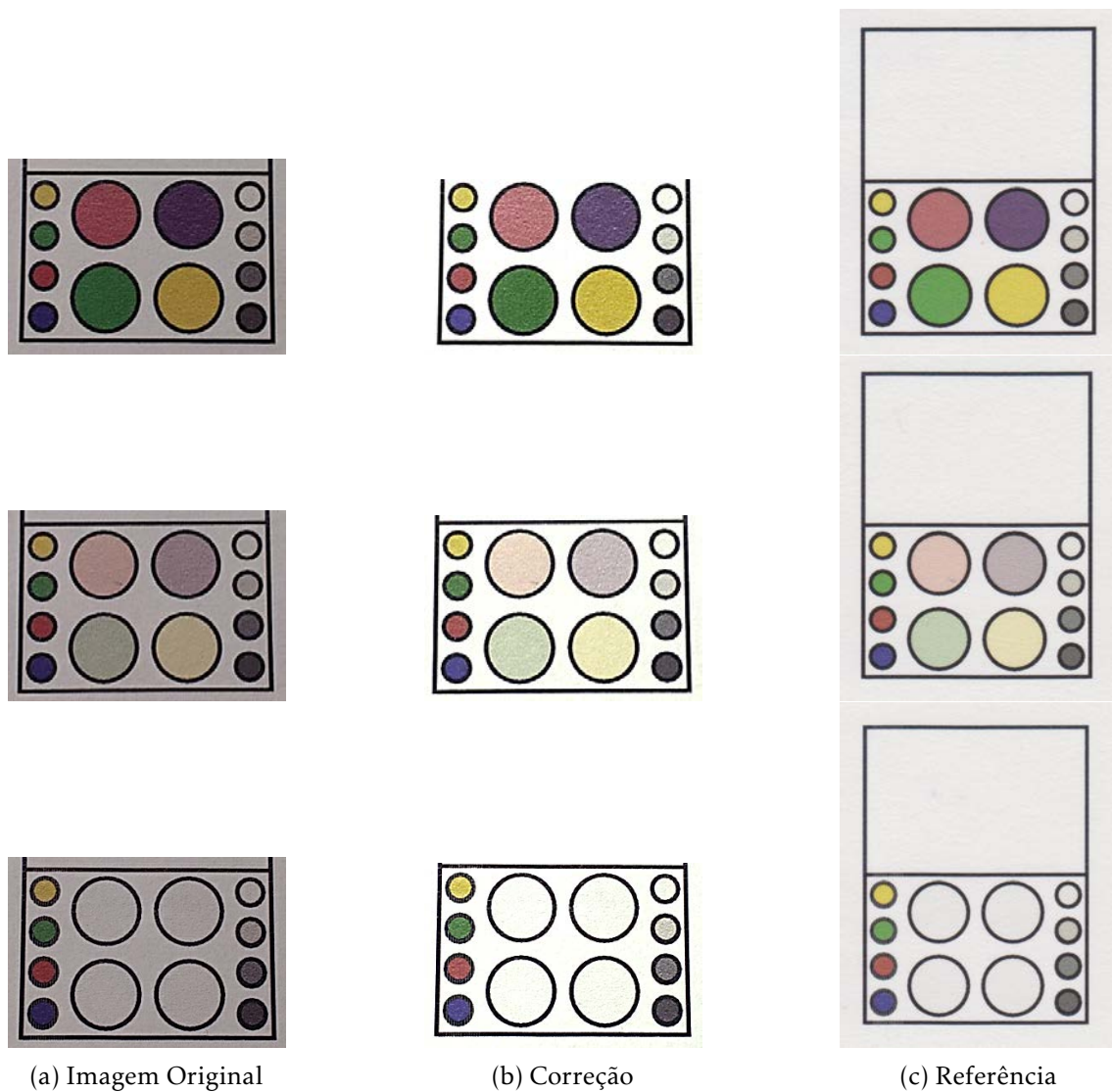


Figura B.17: Correção do sensor com fundo branco (cor intensa, cor média e vazio) fotografado pelo *smartphone Motorola G3* através do algoritmo modificado, descrito na secção 3.2.3.

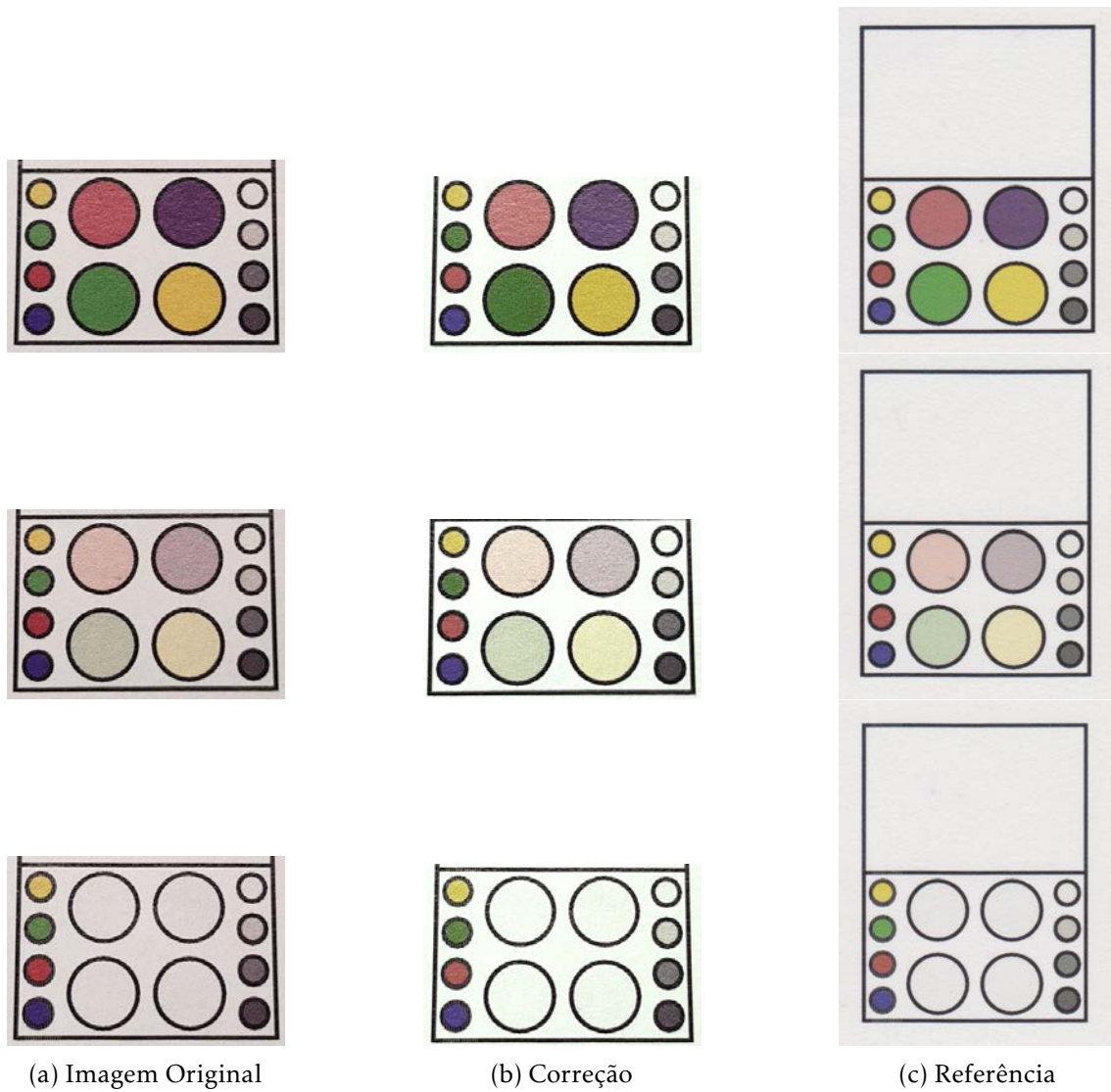


Figura B.18: Correção do sensor com fundo branco (cor intensa, cor média e vazio) fotografado pelo *smartphone Samsung S8* através do algoritmo modificado, descrito na secção 3.2.3.



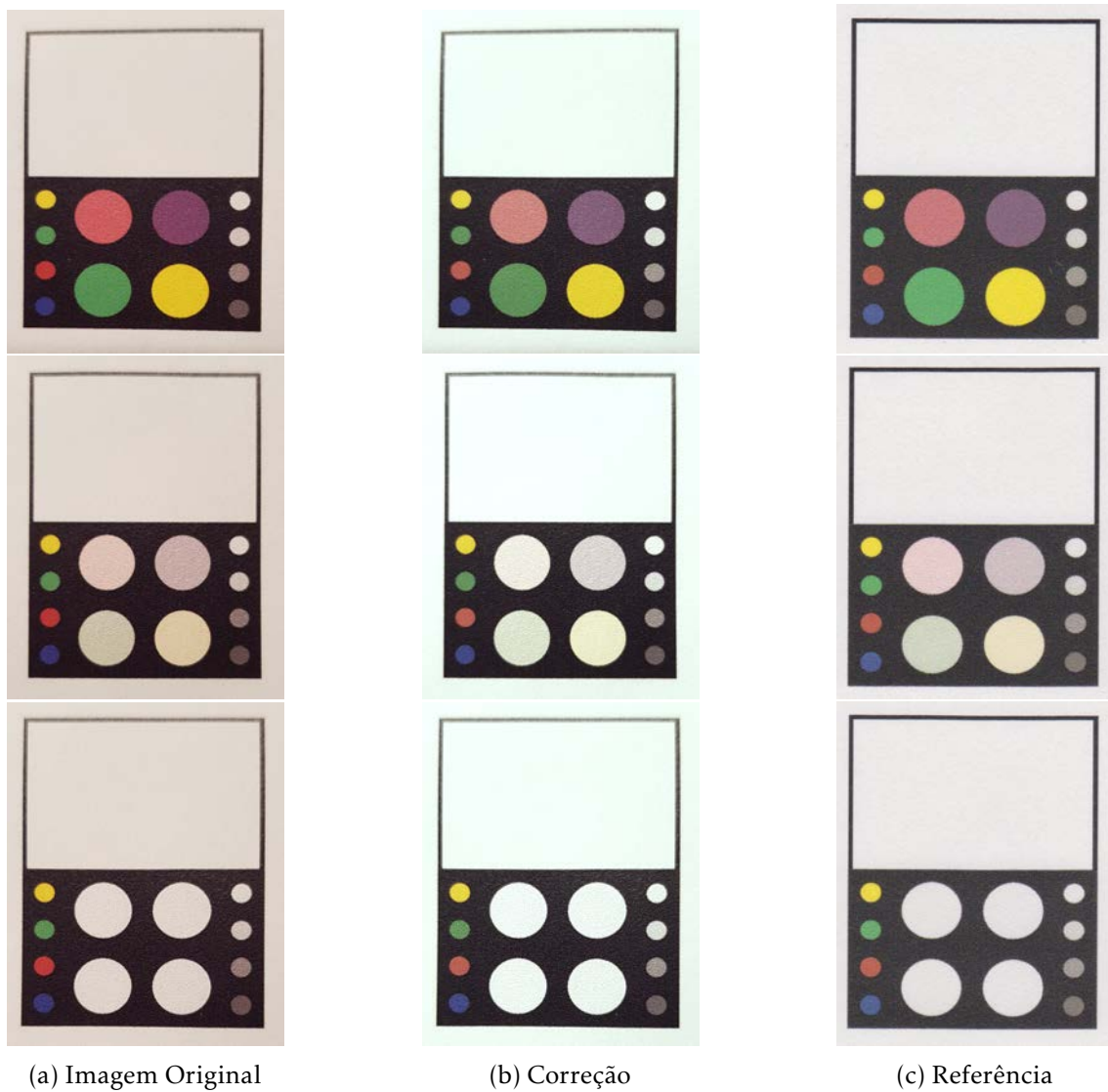


Figura B.19: Correção do sensor com fundo preto (cor intensa, cor média e vazio) fotografado pelo *smartphone Motorola G3* através do algoritmo modificado, descrito na secção 3.2.3.

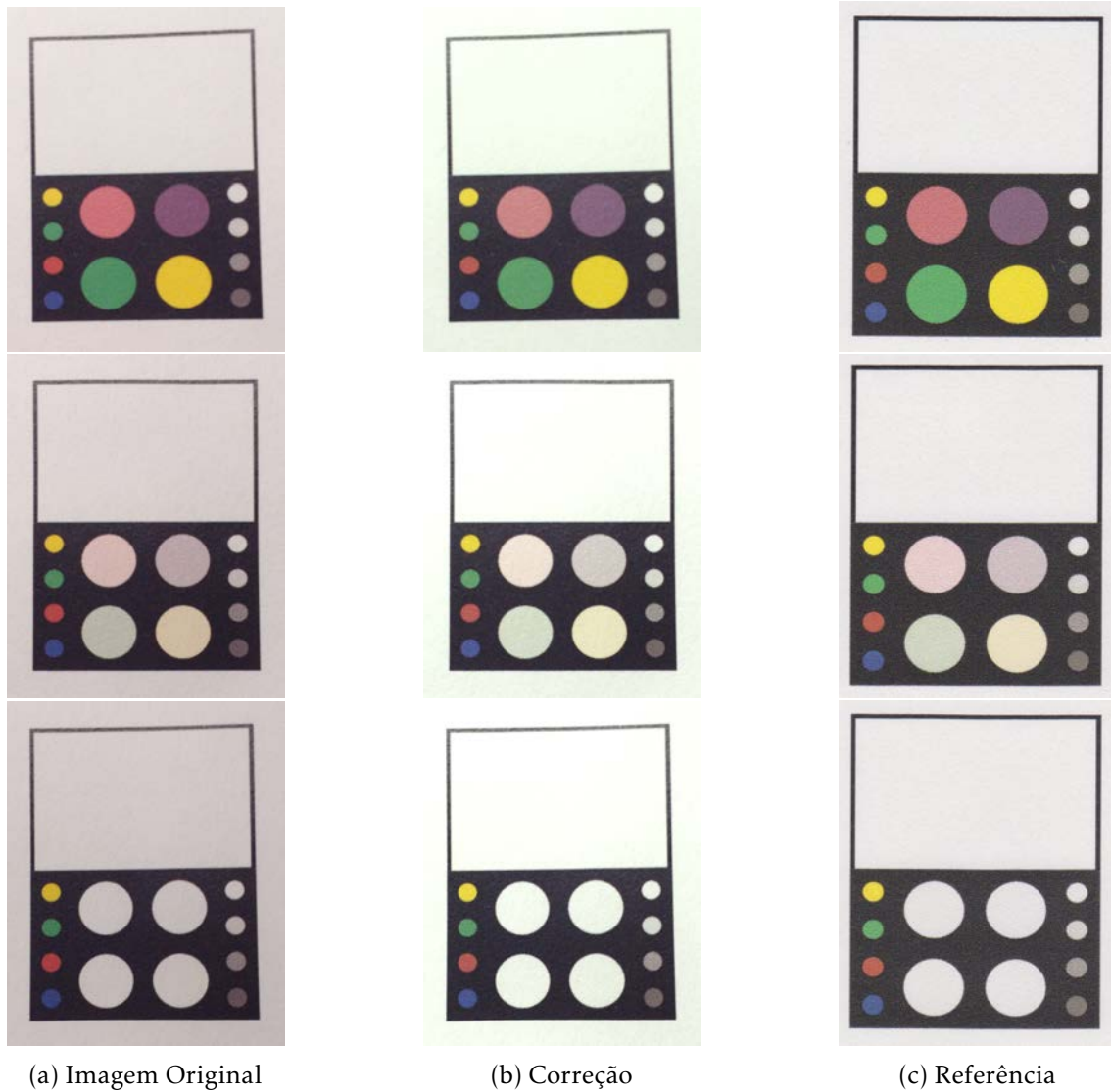


Figura B.20: Correção do sensor com fundo preto (cor intensa, cor média e vazio) fotografado pelo *smartphone Samsung S8* através do algoritmo modificado, descrito na secção 3.2.3.

**B.1.3 Luz Solar**

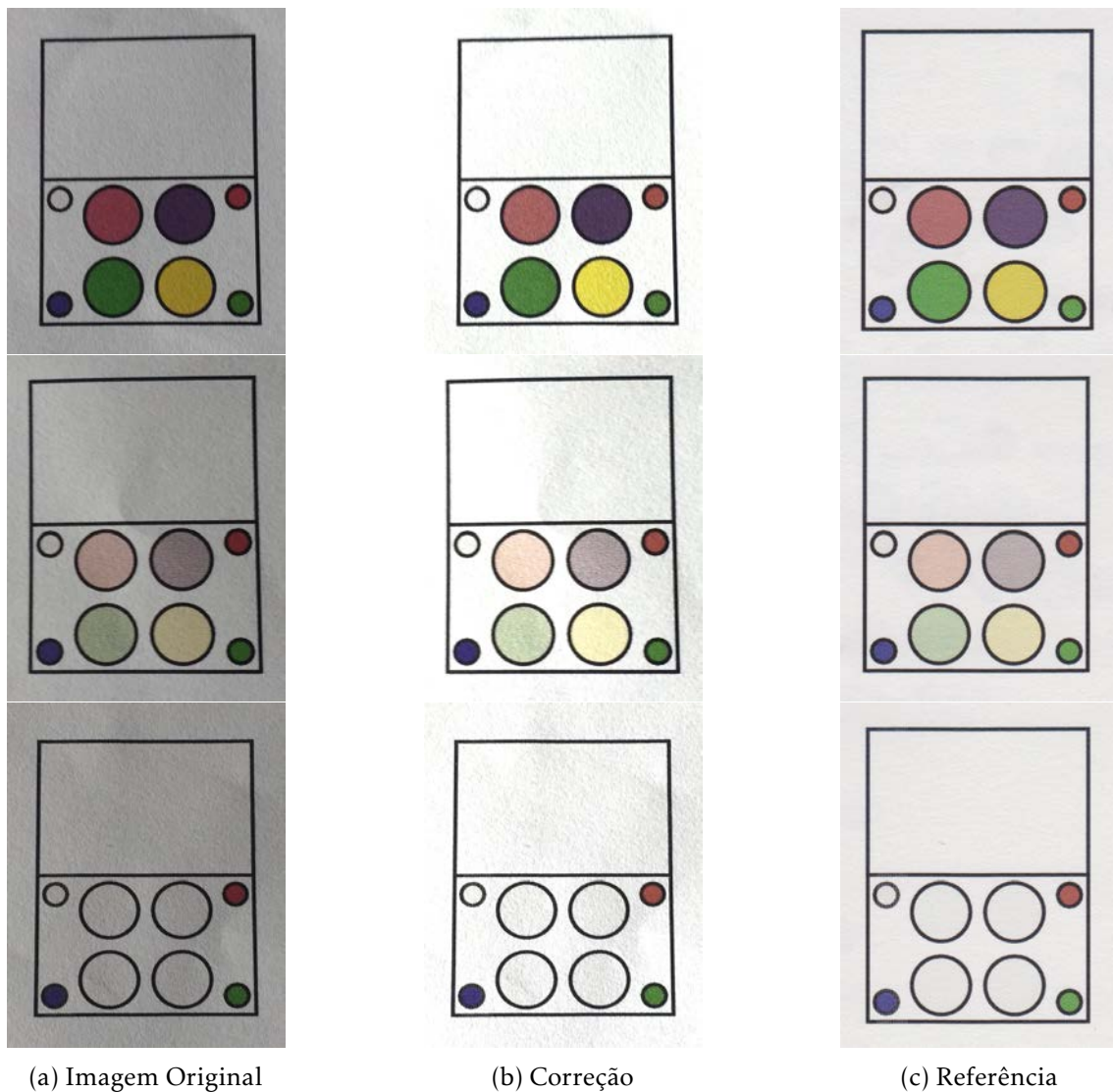


Figura B.21: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Motorola G3* através do algoritmo da versão anterior da aplicação, proposto por Rodrigues [37], descrito na secção 3.2.3.

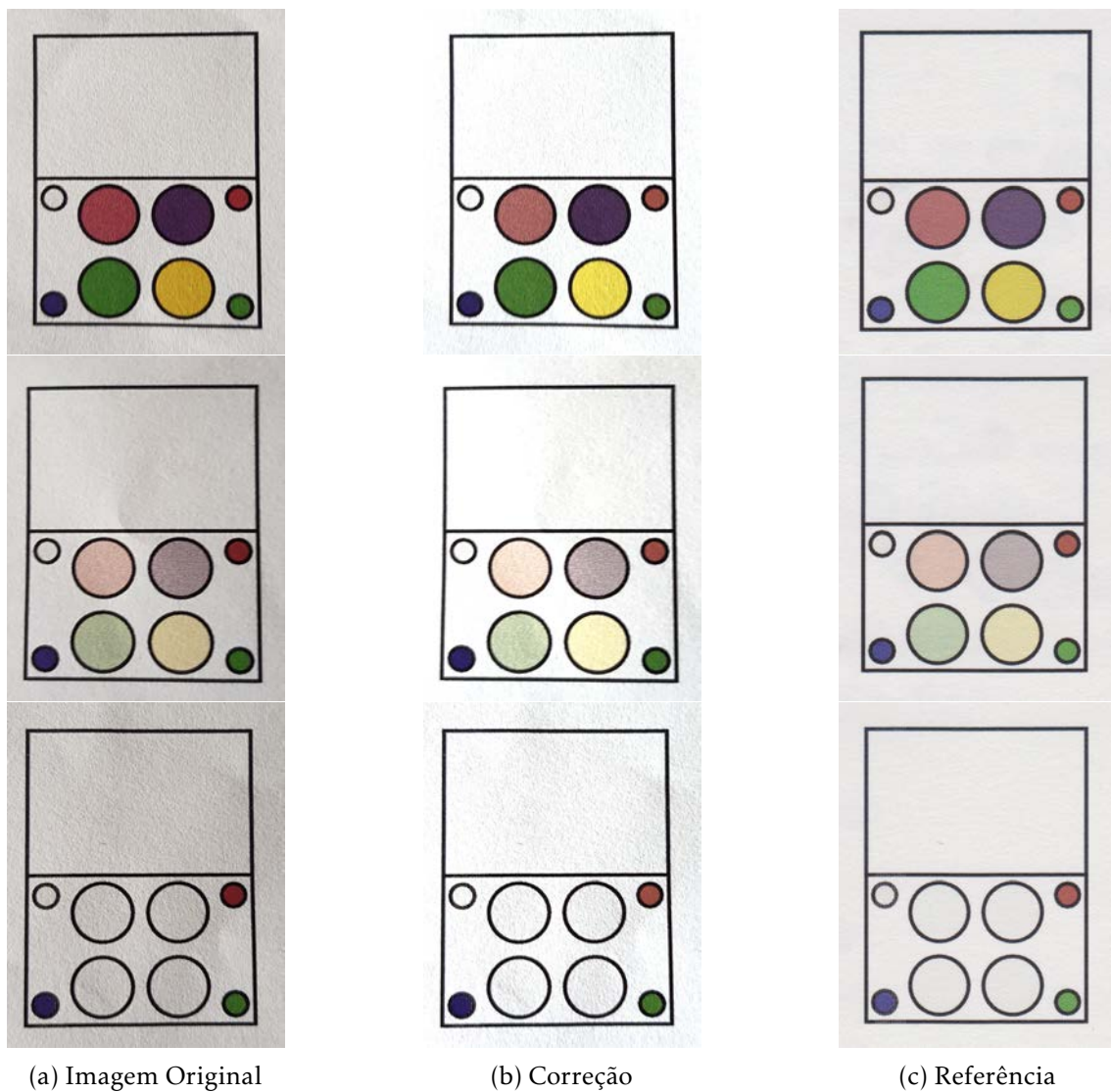


Figura B.22: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Samsung S8* através do algoritmo da versão anterior da aplicação, proposto por Rodrigues [37], descrito na secção 3.2.3.



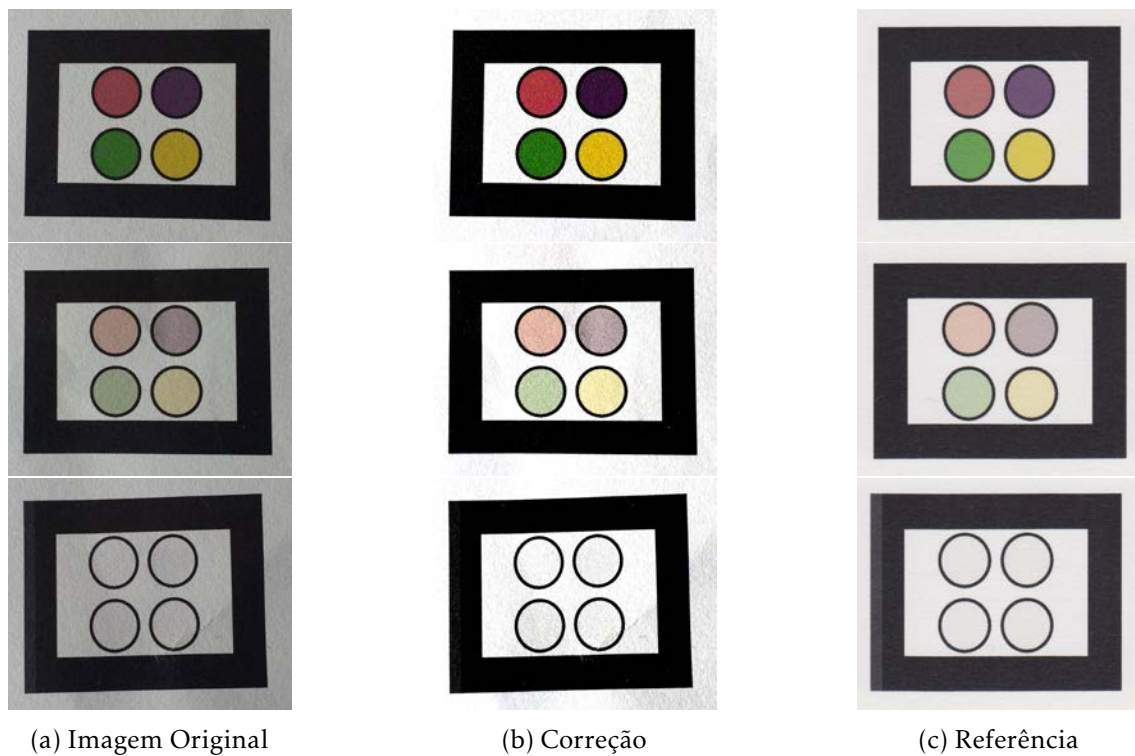


Figura B.23: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Motorola G3* através do algoritmo proposto por Lee e Woo [19], descrito na secção 3.2.2.2.



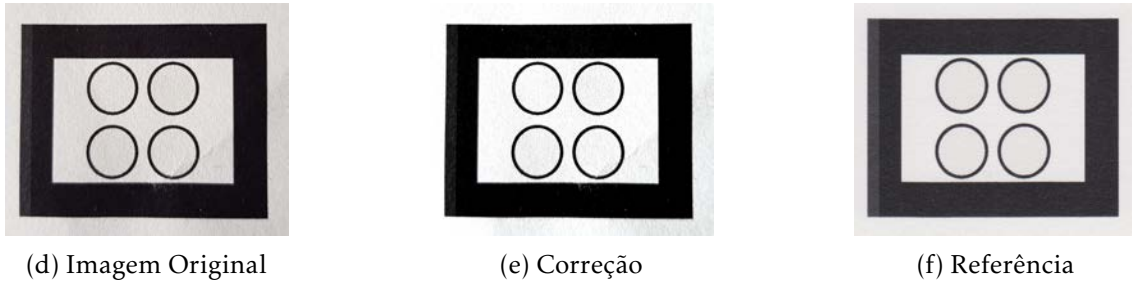


Figura B.23: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Samsung S8* através do algoritmo proposto por Lee e Woo [19], descrito na secção 3.2.2.2.

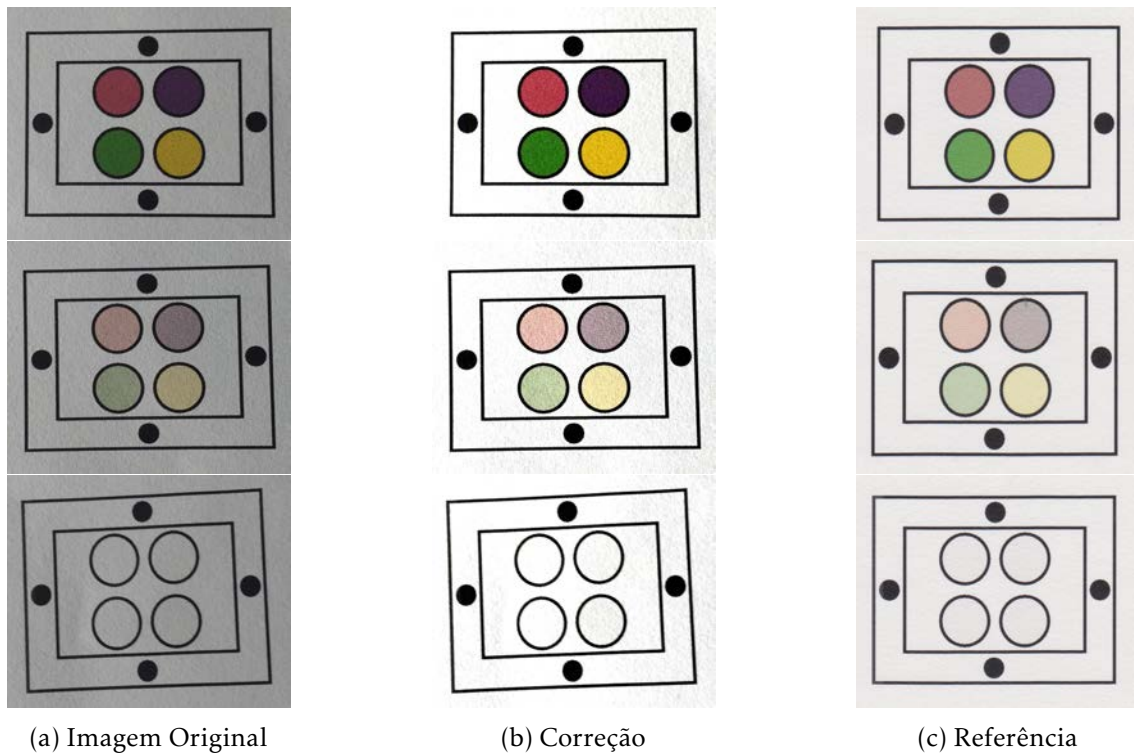


Figura B.24: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Motorola G3* através do algoritmo proposto por Jia et al. [14], descrito em 3.2.2.1.

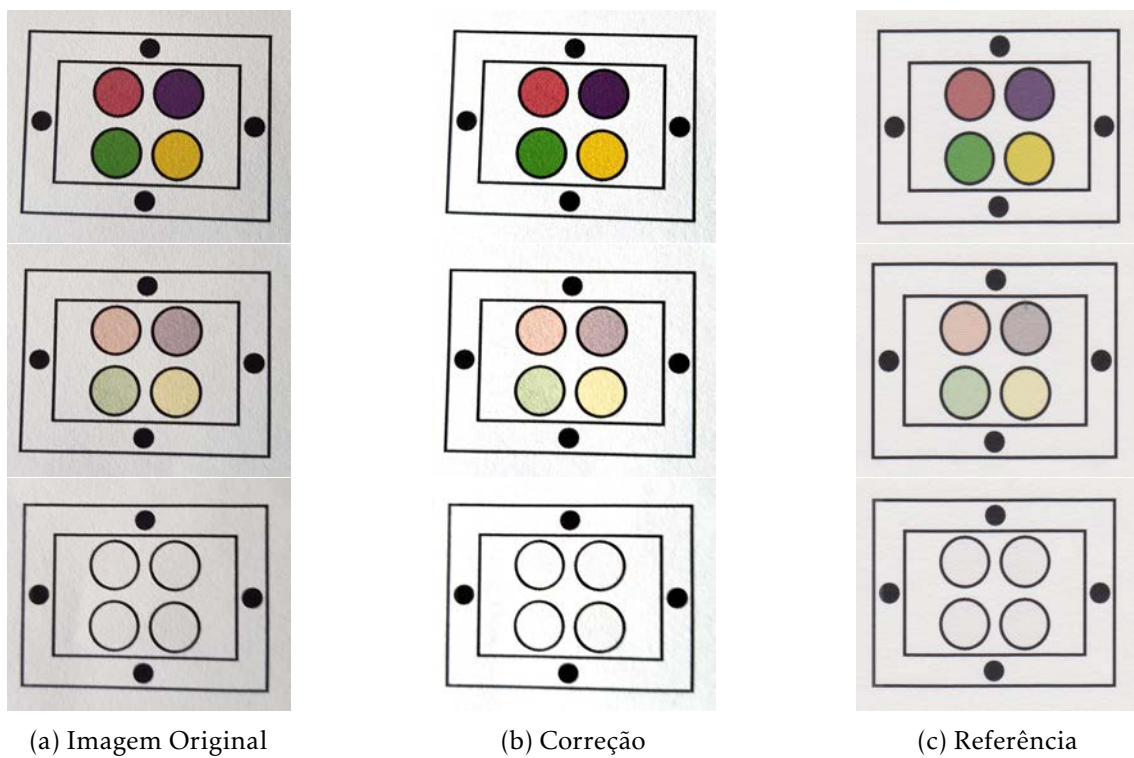


Figura B.25: Correção do sensor (cor intensa, cor média e vazia) fotografado pelo *smartphone Samsung S8* através do algoritmo proposto por Jia et al. [14], descrito em 3.2.2.1.



Figura B.26: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Motorola G3* através do algoritmo proposto por Zhang et al. [54], descrito em 3.2.2.4.





Figura B.27: Correção do sensor (cor intensa, cor média e vazio) fotografado pelo *smartphone Samsung S8* através do algoritmo proposto por Zhang et al. [54], descrito em 3.2.2.4.

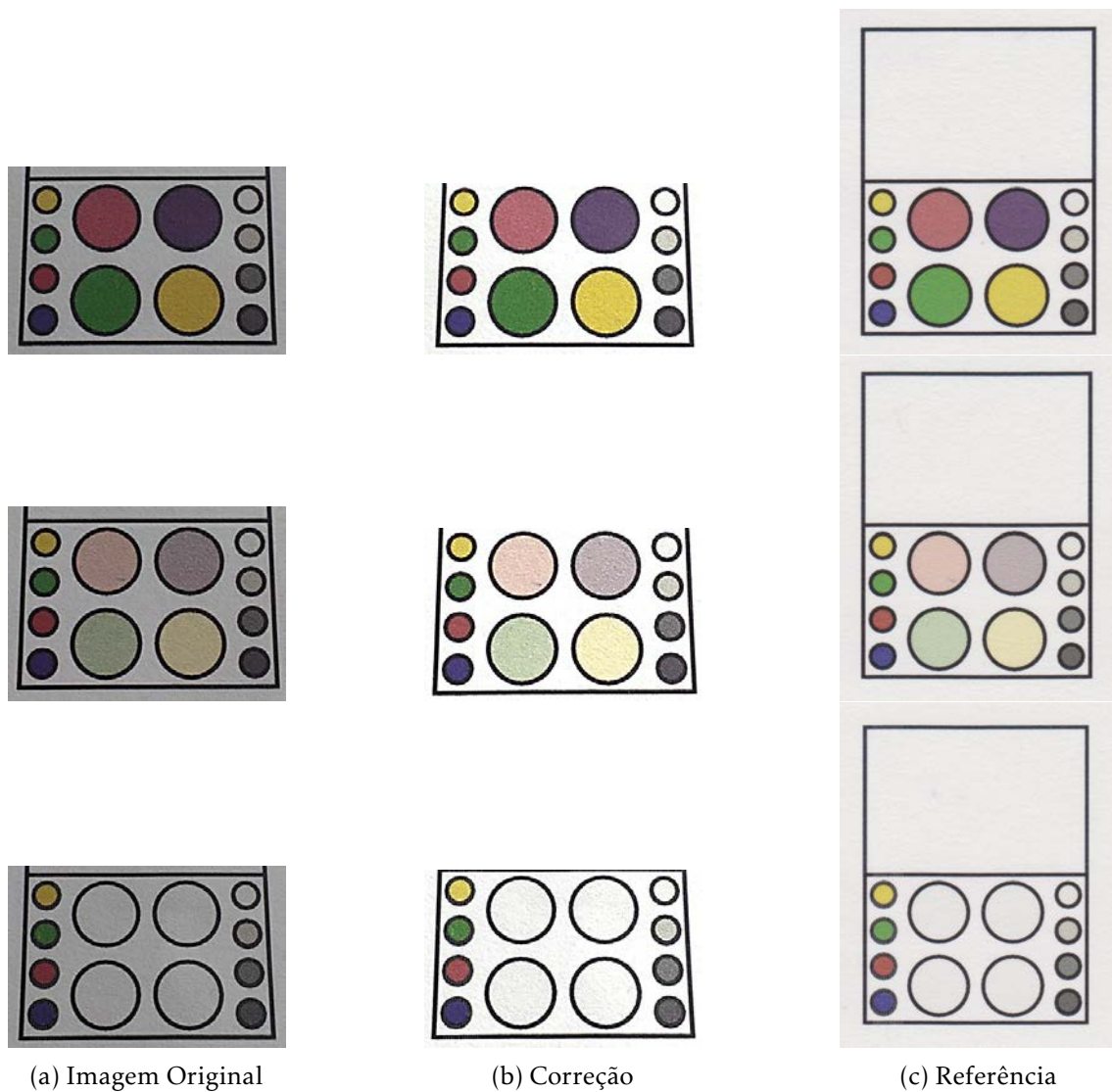


Figura B.28: Correção do sensor com fundo branco (cor intensa, cor média e vazio) fotografado pelo *smartphone Motorola G3* através do algoritmo modificado, descrito na secção 3.2.3.

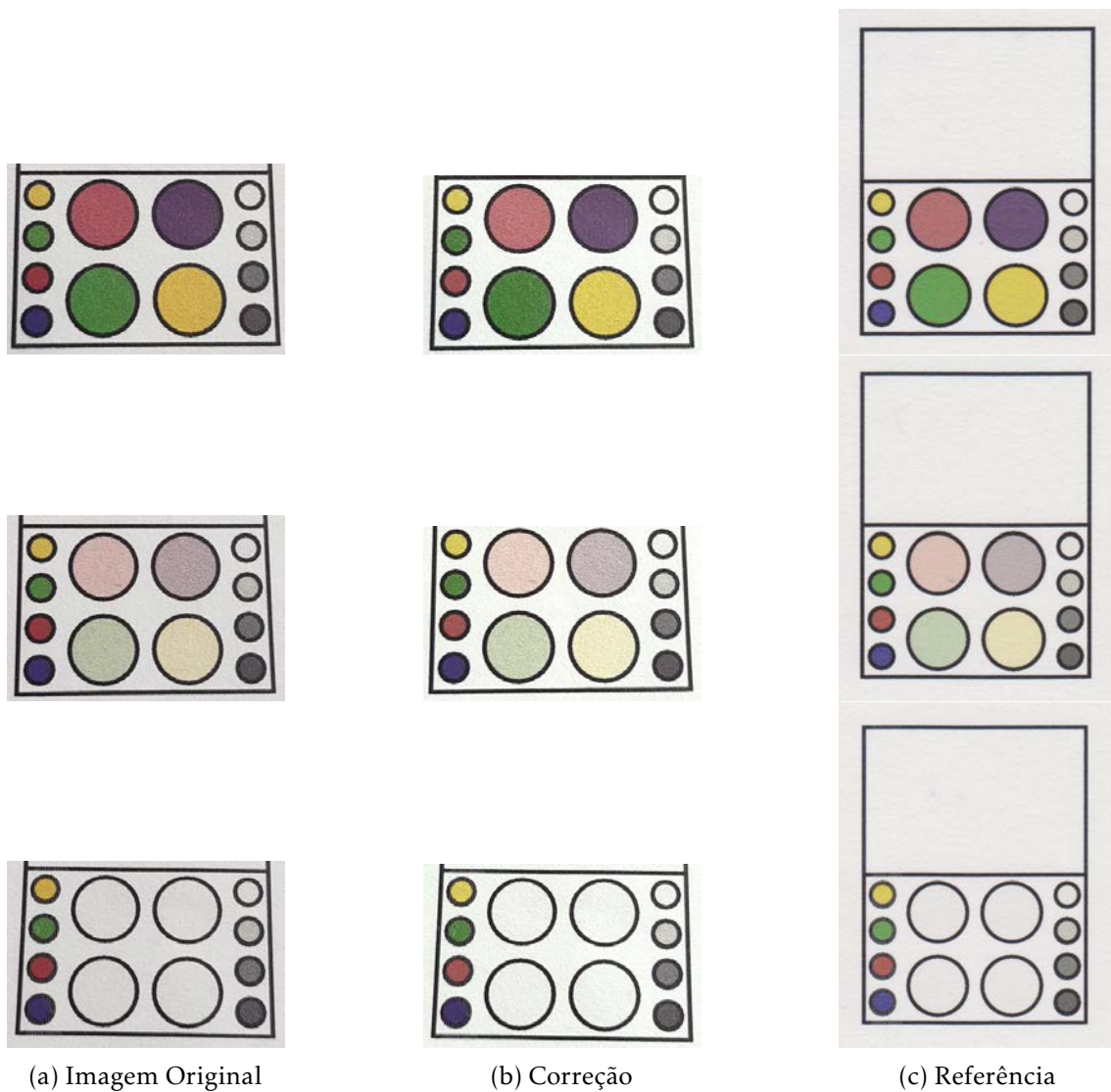


Figura B.29: Correção do sensor com fundo branco (cor intensa, cor média e vazio) fotografado pelo *smartphone Samsung S8* através do algoritmo modificado, descrito na secção 3.2.3.

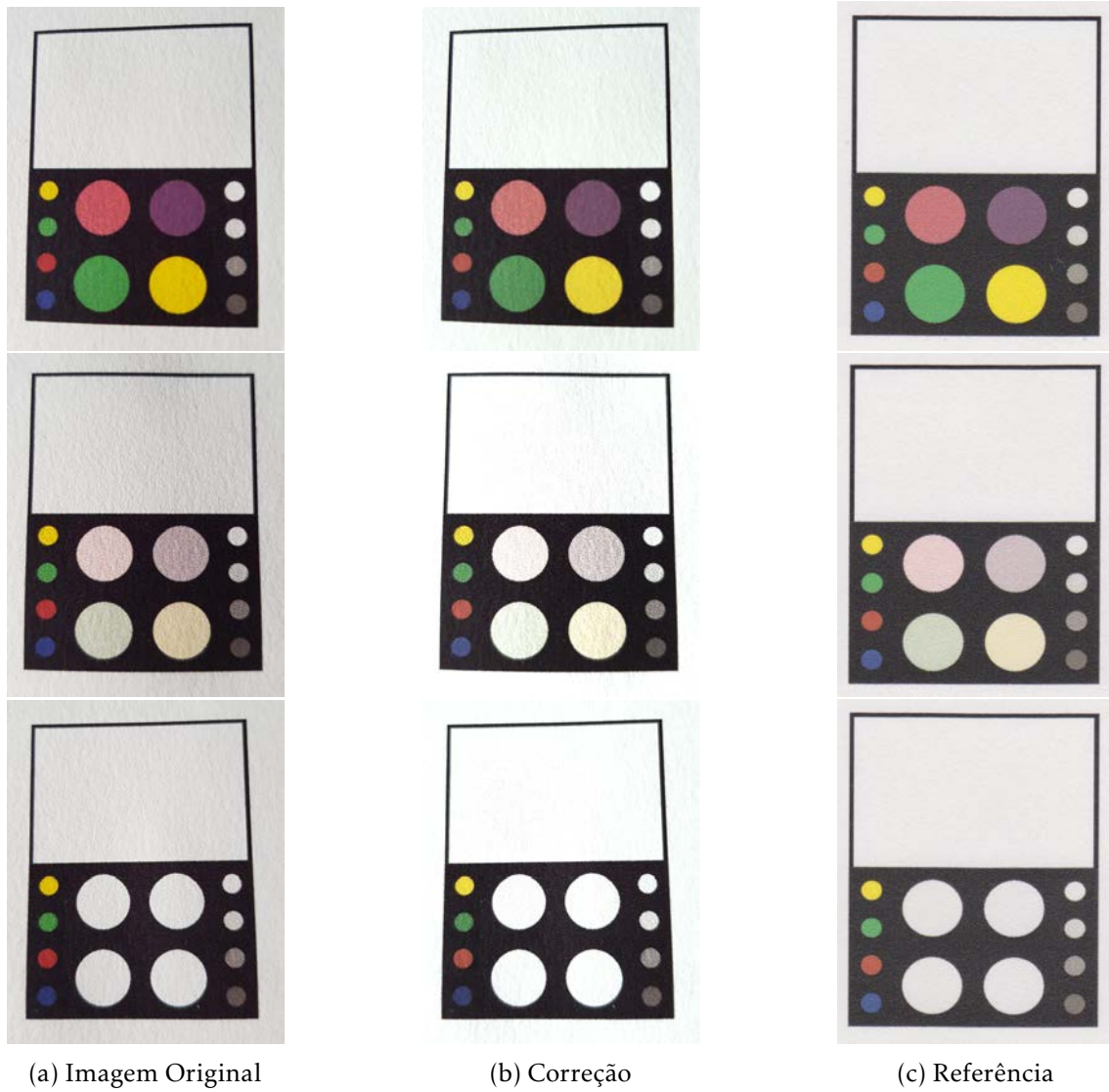


Figura B.30: Correção do sensor com fundo preto (cor intensa, cor média e vazio) fotografado pelo *smartphone Motorola G3* através do algoritmo modificado, descrito na secção 3.2.3.



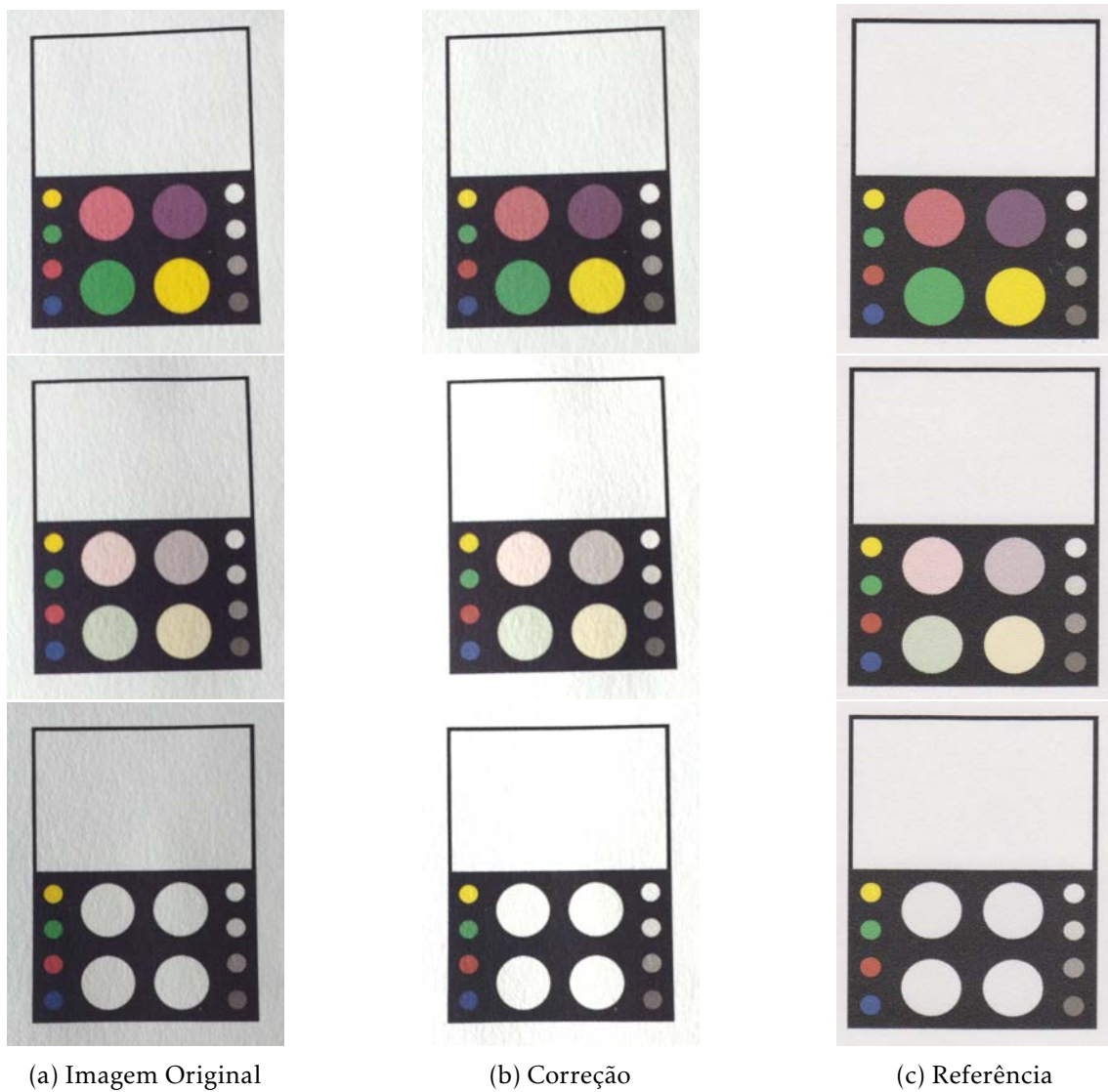


Figura B.31: Correção do sensor com fundo preto (cor intensa, cor média e vazio) fotografado pelo *smartphone Samsung S8* através do algoritmo modificado, descrito na secção 3.2.3.

## B.2 Outros Sensores Usados na Comparação

A implementação dos algoritmos propostos por Lopez-Ruiz et al. e Souza et al. devolve apenas o valor da cor após a sua correção. Assim, não são apresentadas as imagens resultantes mas apenas os *designs* usados para realizar os testes de calibração da cor. Estes algoritmos são descritos na secção 3.2.2.1.

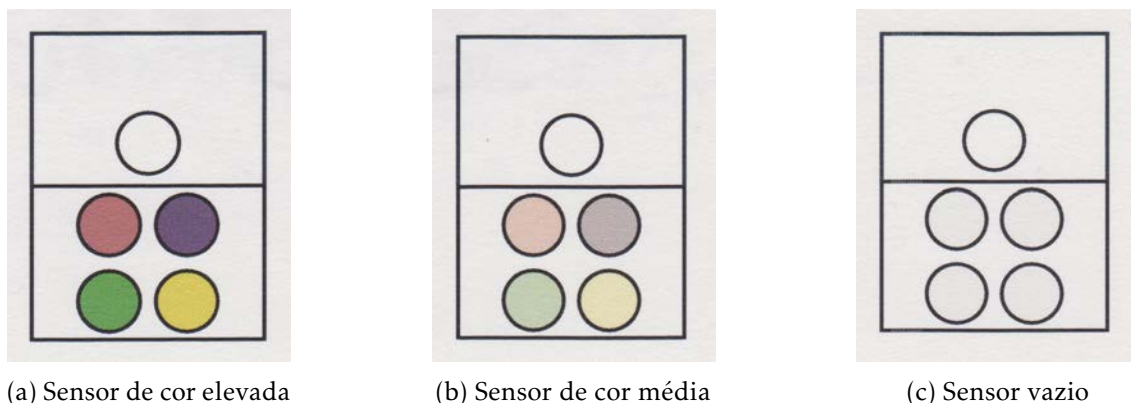
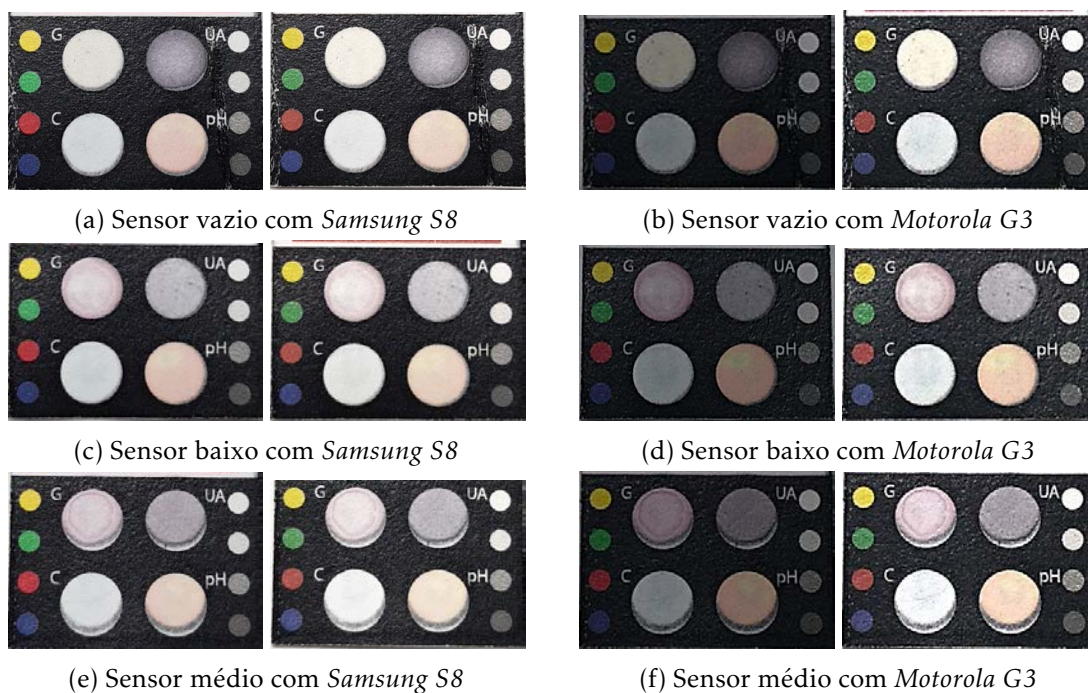
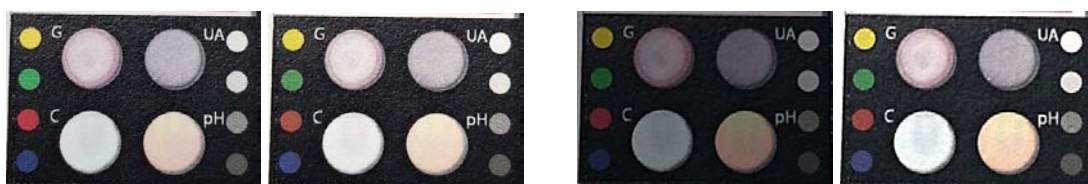


Figura B.32: Sensor usado para testar os algoritmos propostos por Lopez-Ruiz et al. [20] e Souza et al. [44], descritos na secção 3.2.2.1.

## B.3 Resultados da Correção Através da Aplicação Móvel

Após a implementação do método de correção final, os sensores foram corrigidos através da aplicação, produzindo os resultados apresentados nesta secção.

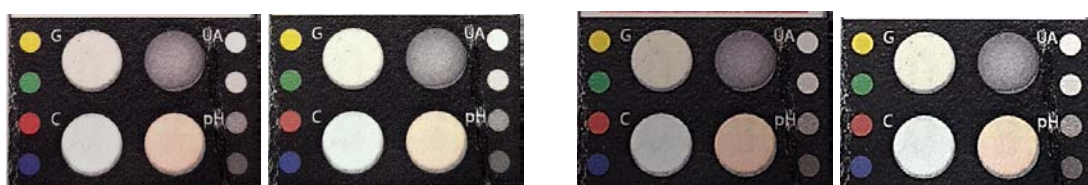




(g) Sensor de alto com *Samsung S8*

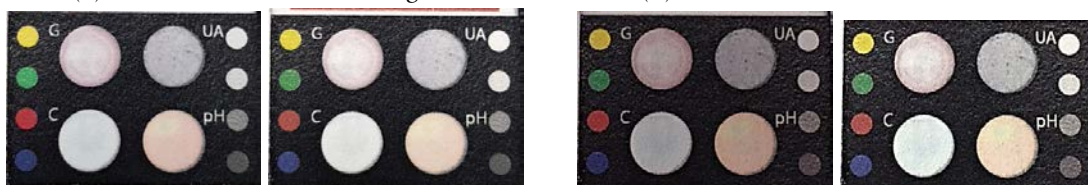
(h) Sensor alto com *Motorola G3*

Figura B.33: Comparação entre o sensor vazio antes e após a correção da cor, através da aplicação. Imagens obtidas com recurso aos dois *smartphones* (*Samsung S8* e *Motorola G3*) sob iluminação fluorescente.



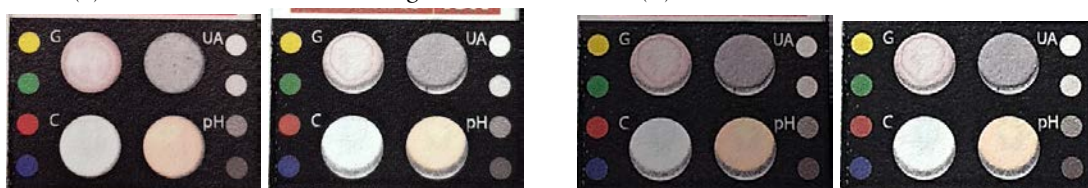
(a) Sensor vazio com *Samsung S8*

(b) Sensor vazio com *Motorola G3*



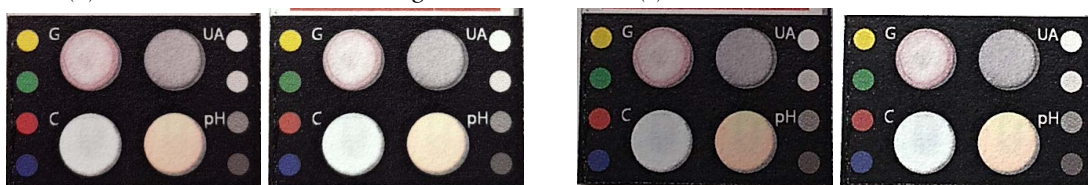
(c) Sensor baixo com *Samsung S8*

(d) Sensor baixo com *Motorola G3*



(e) Sensor médio com *Samsung S8*

(f) Sensor médio com *Motorola G3*



(g) Sensor de alto com *Samsung S8*

(h) Sensor alto com *Motorola G3*

Figura B.34: Comparação entre o sensor vazio antes e após a correção da cor, através da aplicação. Imagens obtidas com recurso aos dois *smartphones* (*Samsung S8* e *Motorola G3*) sob iluminação led.



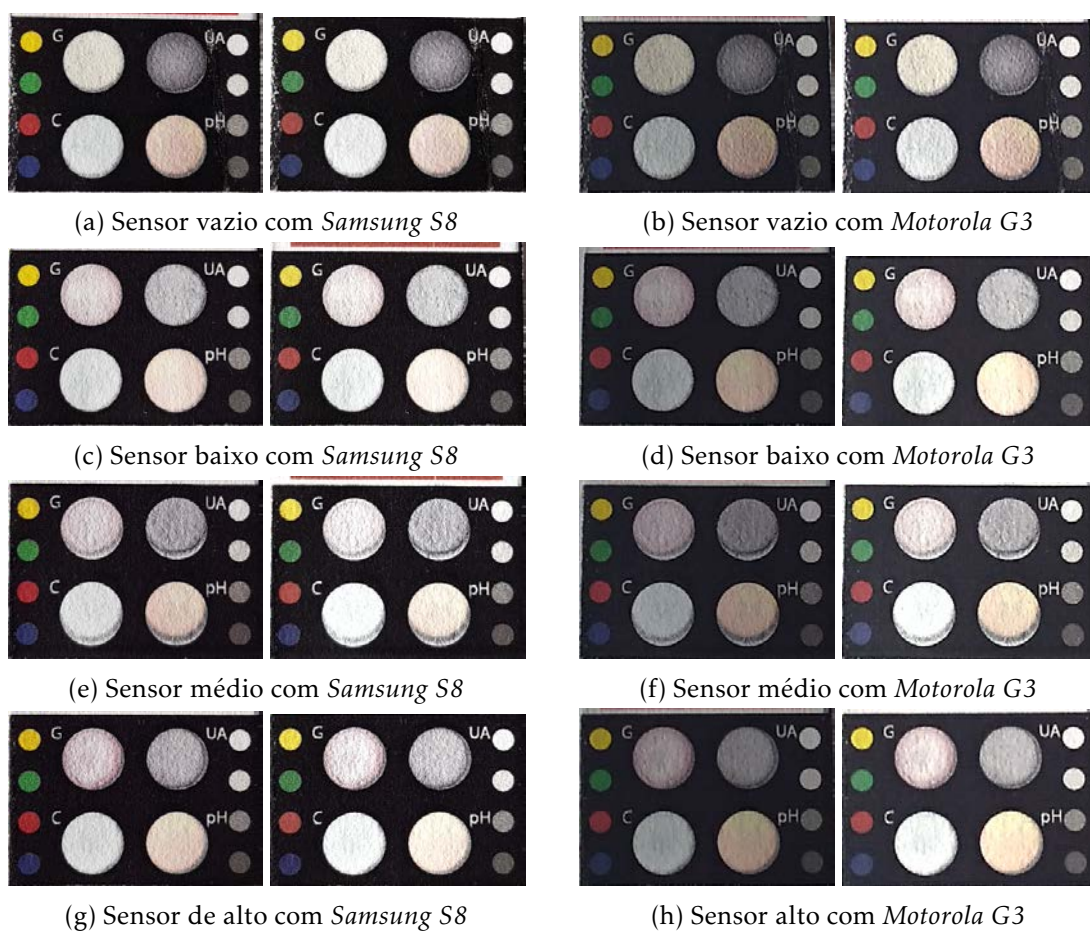


Figura B.35: Comparação entre o sensor vazio antes e após a correção da cor, através da aplicação. Imagens obtidas com recurso aos dois *smartphones* (*Samsung S8* e *Motorola G3*) sob iluminação solar.



## RESULTADOS DA CLASSIFICAÇÃO DO ÁCIDO ÚRICO

Neste apêndice são detalhados os resultados obtidos para os vários modelos testados na classificação das amostras de ácido úrico, cujo processo de análise está descrito na secção 4.1. A discussão dos resultados encontra-se na secção 5.1.2.

Foram considerados três tipos de classificadores (*K-Nearest Neighbours*, *Support Vector Machines* e *Random Forest*) com os quais foram testadas múltiplas combinações de *features* provenientes do trabalho anterior e da análise do próprio conjunto de dados.

Os modelos de classificação das diferentes concentrações e níveis foram explorados usando dados com e sem a aplicação de PCA, sendo avaliados através de três métricas: *F1-Score*, *Recall* e *Precision*.

## C.1 Modelos RGB

Resultados obtidos na classificação das concentrações e dos quatro níveis de ácido úrico, baseados na média das três componentes do modelo de cor RGB.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.41	0.43	0.41	0.36	0.38	0.37
Support Vector Machines	0.39	0.41	0.39	0.36	0.39	0.37
Random Forest	0.38	0.39	0.43	0.42	0.46	0.46
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.35	0.35	0.36	0.34	0.37	0.36
Support Vector Machines	0.47	0.48	0.48	0.40	0.43	0.43
Random Forest	0.21	0.27	0.23	0.24	0.30	0.23

Tabela C.1: Resultados dos modelos criados para a classificação das concentrações do ácido úrico com RGB.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.69	0.69	0.70	0.79	0.79	0.80
Support Vector Machines	0.80	0.80	0.81	0.82	0.82	0.83
Random Forest	0.81	0.81	0.82	0.78	0.78	0.80
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.78	0.79	0.79	0.78	0.78	0.79
Support Vector Machines	0.80	0.80	0.80	0.82	0.82	0.82
Random Forest	0.59	0.61	0.59	0.70	0.70	0.72

Tabela C.2: Resultados dos modelos criados para a classificação dos quatro níveis de ácido úrico com RGB.

## C.2 Modelos HSV

Resultados obtidos na classificação das concentrações e dos quatro níveis de ácido úrico, baseados na média das três componentes do modelo de cor HSV.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.38	0.39	0.41	0.38	0.40	0.42
Support Vector Machines	0.45	0.45	0.50	0.41	0.44	0.43
Random Forest	0.38	0.38	0.41	0.37	0.41	0.39
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.36	0.37	0.36	0.34	0.37	0.36
Support Vector Machines	0.51	0.51	0.51	0.46	0.48	0.50
Random Forest	0.37	0.39	0.36	0.38	0.42	0.39

Tabela C.3: Resultados dos modelos criados para a classificação das concentrações do ácido úrico com HSV.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.81	0.81	0.81	0.79	0.80	0.80
Support Vector Machines	0.82	0.83	0.82	0.80	0.80	0.82
Random Forest	0.78	0.77	0.80	0.80	0.80	0.83
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.78	0.79	0.78	0.79	0.80	0.80
Support Vector Machines	0.80	0.80	0.81	0.80	0.80	0.82
Random Forest	0.82	0.83	0.82	0.76	0.76	0.76

Tabela C.4: Resultados dos modelos criados para a classificação dos quatro níveis de ácido úrico com HSV.

### C.3 Modelos XYZ

Resultados obtidos na classificação das concentrações e dos quatro níveis de ácido úrico, baseados na média das três componentes do modelo de cor XYZ.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.40	0.41	0.40	0.31	0.34	0.31
Support Vector Machines	0.47	0.48	0.48	0.39	0.41	0.40
Random Forest	0.36	0.36	0.38	0.41	0.45	0.42
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.37	0.38	0.38	0.30	0.34	0.31
Support Vector Machines	0.46	0.46	0.47	0.38	0.41	0.39
Random Forest	0.21	0.24	0.20	0.18	0.20	0.19

Tabela C.5: Resultados dos modelos criados para a classificação das concentrações do ácido úrico com XYZ.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.69	0.69	0.69	0.69	0.69	0.72
Support Vector Machines	0.76	0.78	0.77	0.81	0.82	0.82
Random Forest	0.76	0.77	0.76	0.80	0.80	0.81
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.72	0.73	0.72	0.70	0.69	0.71
Support Vector Machines	0.78	0.79	0.79	0.80	0.80	0.82
Random Forest	0.46	0.46	0.46	0.46	0.47	0.49

Tabela C.6: Resultados dos modelos criados para a classificação dos quatro níveis de ácido úrico com XYZ.



## C.4 Modelos HLS

Resultados obtidos na classificação das concentrações e dos quatro níveis de ácido úrico, baseados na média das três componentes do modelo de cor HLS.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.40	0.40	0.40	0.39	0.42	0.40
Support Vector Machines	0.44	0.44	0.44	0.34	0.37	0.36
Random Forest	0.39	0.39	0.42	0.37	0.41	0.38
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.36	0.38	0.36	0.38	0.40	0.41
Support Vector Machines	0.47	0.46	0.47	0.38	0.39	0.39
Random Forest	0.38	0.39	0.39	0.35	0.37	0.36

Tabela C.7: Resultados dos modelos criados para a classificação das concentrações do ácido úrico com HLS.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.81	0.81	0.82	0.80	0.80	0.81
Support Vector Machines	0.83	0.83	0.84	0.81	0.81	0.81
Random Forest	0.78	0.79	0.80	0.80	0.80	0.81
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.77	0.77	0.77	0.81	0.81	0.81
Support Vector Machines	0.85	0.85	0.86	0.82	0.82	0.83
Random Forest	0.76	0.77	0.76	0.75	0.76	0.77

Tabela C.8: Resultados dos modelos criados para a classificação dos quatro níveis de ácido úrico com HLS.

## C.5 Modelos LAB

Resultados obtidos na classificação das concentrações e dos quatro níveis de ácido úrico, baseados na média das três componentes do modelo de cor LAB.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.42	0.43	0.43	0.39	0.42	0.41
Support Vector Machines	0.47	0.47	0.48	0.40	0.43	0.42
Random Forest	0.39	0.39	0.41	0.34	0.37	0.35
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.54	0.53	0.55	0.42	0.45	0.42
Support Vector Machines	0.54	0.54	0.55	0.43	0.47	0.45
Random Forest	0.51	0.53	0.56	0.43	0.46	0.45

Tabela C.9: Resultados dos modelos criados para a classificação das concentrações do ácido úrico com LAB.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.80	0.81	0.80	0.81	0.82	0.82
Support Vector Machines	0.81	0.82	0.81	0.83	0.83	0.84
Random Forest	0.80	0.81	0.82	0.80	0.80	0.81
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.78	0.79	0.78	0.82	0.82	0.82
Support Vector Machines	0.79	0.80	0.79	0.79	0.79	0.79
Random Forest	0.80	0.80	0.80	0.81	0.81	0.82

Tabela C.10: Resultados dos modelos criados para a classificação dos quatro níveis de ácido úrico com LAB.

## C.6 Modelos YUV

Resultados obtidos na classificação das concentrações e dos quatro níveis de ácido úrico, baseados na média das três componentes do modelo de cor YUV.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.47	0.48	0.48	0.41	0.43	0.43
Support Vector Machines	0.48	0.48	0.49	0.39	0.42	0.41
Random Forest	0.37	0.38	0.39	0.36	0.39	0.37
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.50	0.50	0.53	0.40	0.43	0.43
Support Vector Machines	0.51	0.51	0.52	0.40	0.44	0.42
Random Forest	0.43	0.45	0.45	0.37	0.41	0.39

Tabela C.11: Resultados dos modelos criados para a classificação das concentrações do ácido úrico com YUV.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.77	0.78	0.78	0.80	0.81	0.81
Support Vector Machines	0.80	0.80	0.80	0.83	0.84	0.84
Random Forest	0.72	0.73	0.72	0.78	0.78	0.81
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.80	0.80	0.79	0.81	0.82	0.82
Support Vector Machines	0.82	0.83	0.82	0.78	0.78	0.79
Random Forest	0.80	0.80	0.80	0.77	0.78	0.79

Tabela C.12: Resultados dos modelos criados para a classificação dos quatro níveis de ácido úrico com YUV.

## C.7 Rácio R/B

Resultados obtidos na classificação das concentrações e dos quatro níveis de ácido úrico, baseados no rácio entre a média das componentes vermelha e azul do modelo de cor RGB, sem a aplicação de PCA.

Classificador	Conjunto de Teste			Cross-Validation		
	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.30	0.31	0.33	0.40	0.41	0.44
Support Vector Machines	0.28	0.29	0.29	0.39	0.40	0.42
Random Forest	0.28	0.30	0.29	0.38	0.39	0.41

Tabela C.13: Resultados dos modelos criados para a classificação das concentrações do ácido úrico com o rácio entre a média das componentes vermelha e azul do modelo RGB.

Classificador	Conjunto de Teste			Cross-Validation		
	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.69	0.72	0.70	0.71	0.72	0.75
Support Vector Machines	0.67	0.71	0.69	0.69	0.71	0.70
Random Forest	0.66	0.70	0.68	0.70	0.73	0.72

Tabela C.14: Resultados dos modelos criados para a classificação dos quatro níveis de ácido úrico com o rácio entre a média das componentes vermelha e azul do modelo RGB.

## C.8 Componente H do Modelo HSV

Resultados obtidos na classificação das concentrações e dos quatro níveis de ácido úrico, baseados na média da componente H do modelo de cor HSV, sem a aplicação de PCA.

Classificador	Conjunto de Teste			Cross-Validation		
	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.25	0.27	0.24	0.38	0.40	0.40
Support Vector Machines	0.28	0.28	0.29	0.40	0.42	0.41
Random Forest	0.30	0.32	0.40	0.36	0.38	0.38

Tabela C.15: Resultados dos modelos criados para a classificação das concentrações do ácido úrico com a componente H do modelo HSV.

Classificador	Conjunto de Teste			Cross-Validation		
	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.72	0.74	0.72	0.70	0.71	0.74
Support Vector Machines	0.72	0.74	0.72	0.70	0.71	0.74
Random Forest	0.64	0.67	0.65	0.70	0.73	0.72

Tabela C.16: Resultados dos modelos criados para a classificação dos quatro níveis de ácido úrico com a componente H do modelo HSV.

## C.9 Rácio R/B e Componente H do Modelo HSV

Resultados obtidos na classificação das concentrações e dos quatro níveis de ácido úrico, baseados na média da componente H do modelo de cor HSV e no rácio entre a média das componentes vermelha e azul do modelo de cor RGB.

Classificador	Conjunto de Teste			Cross-Validation		
	F1-Score	Recall	Precision	F1-Score	Recall	Precision
Com PCA						
K-Nearest Neighbours	0.28	0.31	0.29	0.39	0.41	0.41
Support Vector Machines	0.27	0.28	0.28	0.41	0.43	0.45
Random Forest	0.27	0.29	0.27	0.38	0.42	0.39
Sem PCA						
K-Nearest Neighbours	0.28	0.31	0.29	0.39	0.41	0.41
Support Vector Machines	0.27	0.28	0.28	0.41	0.43	0.45
Random Forest	0.30	0.31	0.31	0.35	0.38	0.36

Tabela C.17: Resultados dos modelos criados para a classificação das concentrações do ácido úrico com a componente H (HSV) e o rácio entre vermelho e azul (RGB).

Classificador	Conjunto de Teste			Cross-Validation		
	F1-Score	Recall	Precision	F1-Score	Recall	Precision
Com PCA						
K-Nearest Neighbours	0.65	0.68	0.66	0.70	0.72	0.71
Support Vector Machines	0.72	0.74	0.72	0.70	0.71	0.74
Random Forest	0.73	0.73	0.73	0.73	0.73	0.79
Sem PCA						
K-Nearest Neighbours	0.65	0.68	0.66	0.70	0.72	0.71
Support Vector Machines	0.72	0.74	0.72	0.70	0.71	0.74
Random Forest	0.72	0.74	0.72	0.69	0.71	0.71

Tabela C.18: Resultados dos modelos criados para a classificação dos quatro níveis de ácido úrico com a componente H (HSV) e o rácio entre vermelho e azul (RGB).

## C.10 Componentes A e B do Modelo LAB

Resultados obtidos na classificação das concentrações e dos quatro níveis de ácido úrico, baseados na média dos componentes A e B do modelo de cor LAB.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.39	0.39	0.40	0.38	0.42	0.40
Support Vector Machines	0.40	0.42	0.40	0.40	0.42	0.42
Random Forest	0.40	0.41	0.43	0.38	0.40	0.42
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.39	0.39	0.40	0.38	0.42	0.40
Support Vector Machines	0.40	0.42	0.40	0.40	0.42	0.42
Random Forest	0.37	0.40	0.39	0.36	0.40	0.38

Tabela C.19: Resultados dos modelos criados para a classificação das concentrações do ácido úrico com as componentes A e B do modelo LAB.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.64	0.65	0.64	0.74	0.75	0.75
Support Vector Machines	0.64	0.65	0.64	0.76	0.76	0.77
Random Forest	0.69	0.71	0.69	0.74	0.75	0.77
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.64	0.65	0.64	0.74	0.75	0.75
Support Vector Machines	0.64	0.65	0.64	0.76	0.76	0.77
Random Forest	0.69	0.72	0.70	0.75	0.76	0.78

Tabela C.20: Resultados dos modelos criados para a classificação dos quatro níveis de ácido úrico com as componentes A e B do modelo LAB.

## C.11 Componentes H e S do Modelo HSV

Resultados obtidos na classificação das concentrações e dos quatro níveis de ácido úrico, baseados na média dos componentes H e S do modelo de cor HSV.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.35	0.36	0.36	0.44	0.46	0.47
Support Vector Machines	0.47	0.49	0.49	0.42	0.44	0.47
Random Forest	0.41	0.43	0.44	0.39	0.42	0.42
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.35	0.36	0.36	0.44	0.46	0.47
Support Vector Machines	0.47	0.49	0.49	0.42	0.44	0.47
Random Forest	0.40	0.41	0.42	0.37	0.40	0.39

Tabela C.21: Resultados dos modelos criados para a classificação das concentrações do ácido úrico com as componentes H e S do modelo HSV.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.82	0.83	0.82	0.78	0.78	0.79
Support Vector Machines	0.81	0.81	0.81	0.76	0.76	0.79
Random Forest	0.76	0.77	0.77	0.79	0.79	0.80
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.82	0.83	0.82	0.78	0.78	0.79
Support Vector Machines	0.81	0.81	0.81	0.76	0.76	0.79
Random Forest	0.81	0.81	0.81	0.75	0.75	0.77

Tabela C.22: Resultados dos modelos criados para a classificação dos quatro níveis de ácido úrico com as componentes H e S do modelo HSV.

## C.12 Componentes H e S do Modelo HLS

Resultados obtidos na classificação das concentrações e dos quatro níveis de ácido úrico, baseados na média dos componentes H e S do modelo de cor HLS.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.42	0.42	0.42	0.36	0.39	0.37
Support Vector Machines	0.44	0.46	0.46	0.31	0.35	0.31
Random Forest	0.28	0.28	0.31	0.25	0.27	0.28
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.42	0.42	0.42	0.36	0.39	0.37
Support Vector Machines	0.44	0.46	0.46	0.31	0.35	0.31
Random Forest	0.27	0.28	0.28	0.30	0.32	0.31

Tabela C.23: Resultados dos modelos criados para a classificação das concentrações do ácido úrico com as componentes H e S do modelo HLS.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.77	0.79	0.77	0.74	0.75	0.76
Support Vector Machines	0.80	0.80	0.79	0.76	0.76	0.77
Random Forest	0.74	0.75	0.73	0.72	0.73	0.74
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.77	0.79	0.77	0.74	0.75	0.76
Support Vector Machines	0.80	0.80	0.79	0.76	0.76	0.77
Random Forest	0.75	0.76	0.75	0.69	0.70	0.71

Tabela C.24: Resultados dos modelos criados para a classificação dos quatro níveis de ácido úrico com as componentes H e S do modelo HLS.



### C.13 Componentes H e S dos Modelos HSV e HLS

Resultados obtidos na classificação das concentrações e dos quatro níveis de ácido úrico, baseados na média dos componentes H e S dos modelos de cor HSV e HLS.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.39	0.41	0.41	0.41	0.43	0.43
Support Vector Machines	0.43	0.43	0.45	0.38	0.41	0.42
Random Forest	0.41	0.41	0.45	0.32	0.36	0.31
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.46	0.46	0.47	0.42	0.45	0.45
Support Vector Machines	0.49	0.50	0.50	0.44	0.47	0.46
Random Forest	0.34	0.35	0.37	0.35	0.36	0.37

Tabela C.25: Resultados dos modelos criados para a classificação das concentrações do ácido úrico com as componentes H e S dos modelos HSV e HLS.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.79	0.80	0.79	0.79	0.80	0.80
Support Vector Machines	0.81	0.82	0.81	0.81	0.81	0.83
Random Forest	0.77	0.77	0.77	0.80	0.80	0.82
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.82	0.83	0.83	0.80	0.81	0.81
Support Vector Machines	0.85	0.85	0.86	0.80	0.80	0.82
Random Forest	0.81	0.82	0.81	0.75	0.76	0.78

Tabela C.26: Resultados dos modelos criados para a classificação dos quatro níveis de ácido úrico com as componentes H e S dos modelos HSV e HLS.

### C.14 Modelo RGB, Componente V (HSV), Componente Z (XYZ) , Componente A (LAB) e Componente Y (YUV).

Resultados obtidos na classificação das concentrações e dos quatro níveis de ácido úrico, baseados na combinação da média dos componentes do modelo RGB com as componentes V do modelo HSV, Z do modelo XYZ, A do modelo LAB e Y do modelo YUV.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.41	0.43	0.42	0.34	0.37	0.36
Support Vector Machines	0.40	0.41	0.40	0.36	0.39	0.37
Random Forest	0.39	0.39	0.42	0.41	0.44	0.44
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.43	0.44	0.43	0.34	0.36	0.36
Support Vector Machines	0.40	0.42	0.40	0.35	0.38	0.35
Random Forest	0.19	0.23	0.18	0.22	0.26	0.24

Tabela C.27: Resultados dos modelos criados para a classificação das concentrações do ácido úrico com o modelo RGB, e as componentes V (HSV), Z (XYZ) , A (LAB) e Y (YUV).

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.73	0.75	0.74	0.79	0.79	0.79
Support Vector Machines	0.77	0.78	0.79	0.83	0.82	0.84
Random Forest	0.83	0.84	0.83	0.81	0.81	0.83
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.74	0.75	0.75	0.78	0.78	0.79
Support Vector Machines	0.79	0.79	0.80	0.82	0.82	0.83
Random Forest	0.61	0.63	0.61	0.71	0.71	0.73

Tabela C.28: Resultados dos modelos criados para a classificação dos quatro níveis de ácido úrico com o modelo RGB, e as componentes V (HSV), Z (XYZ) , A (LAB) e Y (YUV).

## RESULTADOS DA CLASSIFICAÇÃO DA GLUCOSE

Neste apêndice são detalhados os resultados obtidos para os vários modelos testados para a classificação das amostras de glucose, cujo processo de análise está descrito na secção 4.1. A discussão dos resultados encontra-se na secção 5.1.1.4.

Foram considerados três tipos de classificadores (*K-Nearest Neighbours*, *Support Vector Machines* e *Random Forest*) com os quais foram testadas múltiplas combinações de *features* provenientes do trabalho anterior e da análise do próprio conjunto de dados.

Os modelos de classificação das diferentes concentrações e níveis foram explorados usando dados com e sem a aplicação de PCA, sendo avaliados através de três métricas: *F1-Score*, *Recall* e *Precision*.

## D.1 Modelos Com a Média de RGB

Resultados obtidos na classificação das concentrações e dos níveis de glucose, nos estados de jejum e duas horas após a refeição, baseados na média das três componentes do modelo de cor RGB.

### D.1.1 Concentrações

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.40	0.41	0.42	0.40	0.43	0.39
Support Vector Machines	0.50	0.50	0.55	0.47	0.49	0.48
Random Forest	0.44	0.46	0.44	0.40	0.43	0.40
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.54	0.53	0.58	0.45	0.48	0.46
Support Vector Machines	0.60	0.59	0.71	0.50	0.53	0.51
Random Forest	0.40	0.42	0.48	0.36	0.41	0.34

Tabela D.1: Resultados dos modelos criados para a classificação das concentrações da glucose com a média das componentes do modelo RGB.

### D.1.2 Cinco Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.64	0.64	0.65	0.66	0.66	0.67
Support Vector Machines	0.70	0.70	0.71	0.66	0.66	0.68
Random Forest	0.65	0.64	0.67	0.65	0.65	0.67
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.68	0.68	0.69	0.67	0.67	0.69
Support Vector Machines	0.72	0.72	0.73	0.71	0.71	0.73
Random Forest	0.55	0.57	0.56	0.61	0.62	0.63

Tabela D.2: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com a média das componentes do modelo RGB.

## D.1. MODELOS COM A MÉDIA DE RGB

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.68	0.66	0.71	0.68	0.68	0.71
Support Vector Machines	0.69	0.67	0.72	0.68	0.67	0.73
Random Forest	0.69	0.67	0.73	0.67	0.66	0.73
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.72	0.70	0.75	0.74	0.74	0.77
Support Vector Machines	0.74	0.72	0.76	0.71	0.70	0.73
Random Forest	0.65	0.64	0.71	0.65	0.63	0.72

Tabela D.3: Resultados dos modelos criados para a classificação dos cinco níveis de glucose, duas horas após a refeição com a média das componentes do modelo RGB.

### D.1.3 Quatro Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.73	0.73	0.74	0.71	0.72	0.72
Support Vector Machines	0.70	0.70	0.70	0.73	0.73	0.75
Random Forest	0.70	0.70	0.71	0.72	0.73	0.73
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.76	0.76	0.76	0.74	0.75	0.75
Support Vector Machines	0.75	0.74	0.75	0.75	0.75	0.76
Random Forest	0.73	0.72	0.76	0.70	0.71	0.73

Tabela D.4: Resultados dos modelos criados para a classificação dos quatro níveis de glucose em jejum com a média das componentes do modelo RGB.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.82	0.82	0.83	0.83	0.82	0.86
Support Vector Machines	0.83	0.84	0.83	0.84	0.83	0.87
Random Forest	0.86	0.84	0.88	0.84	0.81	0.90
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.84	0.83	0.86	0.84	0.82	0.87
Support Vector Machines	0.86	0.86	0.87	0.83	0.82	0.86
Random Forest	0.81	0.80	0.82	0.79	0.77	0.85

Tabela D.5: Resultados dos modelos criados para a classificação dos quatro níveis de glucose, duas horas após a refeição com a média das componentes do modelo RGB.

## D.2 Modelos Com a Média de HSV

Resultados obtidos na classificação das concentrações e dos níveis de glucose, nos estados de jejum e duas horas após a refeição, baseados na média das três componentes do modelo de cor HSV.

### D.2.1 Concentrações

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.51	0.51	0.54	0.47	0.49	0.48
Support Vector Machines	0.50	0.50	0.64	0.49	0.51	0.50
Random Forest	0.37	0.43	0.35	0.39	0.44	0.38
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.54	0.54	0.54	0.51	0.53	0.53
Support Vector Machines	0.56	0.55	0.60	0.52	0.54	0.54
Random Forest	0.41	0.46	0.39	0.40	0.45	0.37

Tabela D.6: Resultados dos modelos criados para a classificação das concentrações da glucose com a média das componentes do modelo HSV.

### D.2.2 Cinco Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.64	0.64	0.65	0.66	0.66	0.69
Support Vector Machines	0.70	0.70	0.71	0.66	0.66	0.68
Random Forest	0.67	0.67	0.68	0.64	0.64	0.65
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.62	0.62	0.64	0.67	0.67	0.69
Support Vector Machines	0.66	0.66	0.67	0.70	0.70	0.72
Random Forest	0.66	0.66	0.67	0.67	0.68	0.69

Tabela D.7: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com a média das componentes do modelo HSV.

## D.2. MODELOS COM A MÉDIA DE HSV

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.70	0.68	0.75	0.70	0.69	0.75
Support Vector Machines	0.67	0.66	0.78	0.67	0.66	0.77
Random Forest	0.70	0.68	0.75	0.68	0.67	0.73
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.73	0.71	0.75	0.73	0.72	0.77
Support Vector Machines	0.69	0.68	0.83	0.69	0.69	0.77
Random Forest	0.71	0.70	0.75	0.71	0.71	0.73

Tabela D.8: Resultados dos modelos criados para a classificação dos cinco níveis de glucose, duas horas após a refeição com a média das componentes do modelo HSV.

### D.2.3 Quatro Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.75	0.75	0.74	0.72	0.73	0.73
Support Vector Machines	0.76	0.75	0.81	0.71	0.72	0.79
Random Forest	0.72	0.72	0.72	0.73	0.74	0.74
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.75	0.75	0.75	0.75	0.75	0.76
Support Vector Machines	0.76	0.75	0.78	0.72	0.72	0.76
Random Forest	0.76	0.76	0.77	0.75	0.76	0.77

Tabela D.9: Resultados dos modelos criados para a classificação dos quatro níveis de glucose em jejum com a média das componentes do modelo HSV.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.81	0.81	0.81	0.85	0.84	0.88
Support Vector Machines	0.86	0.85	0.87	0.84	0.82	0.88
Random Forest	0.85	0.83	0.86	0.82	0.80	0.88
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.86	0.85	0.87	0.87	0.85	0.90
Support Vector Machines	0.82	0.81	0.84	0.82	0.80	0.86
Random Forest	0.84	0.86	0.83	0.83	0.82	0.83

Tabela D.10: Resultados dos modelos criados para a classificação dos quatro níveis de glucose, duas horas após a refeição com a média das componentes do modelo HSV.

### D.3 Modelos Com a Média de XYZ

Resultados obtidos na classificação das concentrações e dos níveis de glucose, nos estados de jejum e duas horas após a refeição, baseados na média das três componentes do modelo de cor XYZ.

#### D.3.1 Concentrações

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.44	0.44	0.47	0.41	0.43	0.41
Support Vector Machines	0.48	0.49	0.47	0.44	0.47	0.44
Random Forest	0.39	0.41	0.39	0.42	0.45	0.41
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.45	0.45	0.47	0.41	0.43	0.42
Support Vector Machines	0.61	0.60	0.63	0.50	0.54	0.51
Random Forest	0.29	0.35	0.27	0.33	0.38	0.31

Tabela D.11: Resultados dos modelos criados para a classificação das concentrações da glucose com a média das componentes do modelo XYZ.

#### D.3.2 Cinco Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.60	0.59	0.60	0.60	0.61	0.61
Support Vector Machines	0.66	0.67	0.66	0.68	0.68	0.70
Random Forest	0.65	0.66	0.65	0.67	0.67	0.69
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.64	0.63	0.65	0.65	0.65	0.66
Support Vector Machines	0.72	0.72	0.72	0.72	0.72	0.73
Random Forest	0.50	0.52	0.56	0.52	0.54	0.55

Tabela D.12: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com a média das componentes do modelo XYZ.



### D.3. MODELOS COM A MÉDIA DE XYZ

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.66	0.64	0.69	0.67	0.66	0.72
Support Vector Machines	0.67	0.65	0.74	0.67	0.66	0.75
Random Forest	0.69	0.67	0.73	0.70	0.68	0.77
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.70	0.68	0.73	0.70	0.69	0.74
Support Vector Machines	0.76	0.74	0.80	0.74	0.74	0.77
Random Forest	0.65	0.63	0.73	0.63	0.62	0.70

Tabela D.13: Resultados dos modelos criados para a classificação dos cinco níveis de glucose, duas horas após a refeição com a média das componentes do modelo XYZ.

#### D.3.3 Quatro Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.69	0.70	0.69	0.67	0.68	0.69
Support Vector Machines	0.71	0.71	0.76	0.70	0.71	0.78
Random Forest	0.70	0.70	0.72	0.72	0.73	0.73
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.75	0.75	0.75	0.71	0.71	0.72
Support Vector Machines	0.78	0.78	0.78	0.77	0.77	0.79
Random Forest	0.68	0.68	0.74	0.63	0.64	0.65

Tabela D.14: Resultados dos modelos criados para a classificação dos quatro níveis de glucose em jejum com a média das componentes do modelo XYZ.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.82	0.80	0.85	0.78	0.76	0.83
Support Vector Machines	0.85	0.82	0.89	0.79	0.76	0.88
Random Forest	0.84	0.82	0.88	0.81	0.78	0.89
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.83	0.81	0.87	0.79	0.78	0.85
Support Vector Machines	0.87	0.87	0.87	0.82	0.81	0.84
Random Forest	0.84	0.81	0.88	0.76	0.74	0.86

Tabela D.15: Resultados dos modelos criados para a classificação dos quatro níveis de glucose, duas horas após a refeição com a média das componentes do modelo XYZ.

## D.4 Modelos Com a Média de HLS

Resultados obtidos na classificação das concentrações e dos níveis de glucose, nos estados de jejum e duas horas após a refeição, baseados na média das três componentes do modelo de cor HLS.

### D.4.1 Concentrações

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.52	0.53	0.54	0.47	0.49	0.49
Support Vector Machines	0.49	0.49	0.59	0.51	0.52	0.53
Random Forest	0.42	0.46	0.45	0.40	0.46	0.38
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.54	0.54	0.59	0.53	0.56	0.55
Support Vector Machines	0.53	0.52	0.57	0.53	0.54	0.55
Random Forest	0.40	0.44	0.41	0.43	0.46	0.42

Tabela D.16: Resultados dos modelos criados para a classificação das concentrações da glucose com a média das componentes do modelo HLS.

### D.4.2 Cinco Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.68	0.68	0.68	0.66	0.66	0.67
Support Vector Machines	0.64	0.63	0.67	0.67	0.67	0.69
Random Forest	0.67	0.67	0.71	0.66	0.66	0.69
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.67	0.67	0.67	0.68	0.68	0.69
Support Vector Machines	0.66	0.65	0.67	0.70	0.70	0.72
Random Forest	0.69	0.69	0.71	0.68	0.68	0.70

Tabela D.17: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com a média das componentes do modelo HLS.

#### D.4. MODELOS COM A MÉDIA DE HLS

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.68	0.66	0.72	0.69	0.68	0.71
Support Vector Machines	0.61	0.59	0.66	0.68	0.66	0.72
Random Forest	0.70	0.71	0.72	0.67	0.67	0.71
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.71	0.69	0.75	0.73	0.73	0.75
Support Vector Machines	0.63	0.61	0.67	0.70	0.69	0.74
Random Forest	0.71	0.70	0.74	0.70	0.68	0.75

Tabela D.18: Resultados dos modelos criados para a classificação dos cinco níveis de glucose, duas horas após a refeição com a média das componentes do modelo HLS.

#### D.4.3 Quatro Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.74	0.74	0.76	0.71	0.72	0.72
Support Vector Machines	0.75	0.74	0.77	0.72	0.73	0.76
Random Forest	0.75	0.75	0.76	0.72	0.72	0.73
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.74	0.74	0.74	0.73	0.74	0.74
Support Vector Machines	0.74	0.73	0.75	0.74	0.74	0.76
Random Forest	0.71	0.71	0.72	0.71	0.72	0.72

Tabela D.19: Resultados dos modelos criados para a classificação dos quatro níveis de glucose em jejum com a média das componentes do modelo HLS.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.86	0.85	0.88	0.86	0.84	0.90
Support Vector Machines	0.83	0.83	0.83	0.84	0.83	0.87
Random Forest	0.85	0.87	0.83	0.85	0.86	0.84
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.86	0.86	0.87	0.86	0.85	0.89
Support Vector Machines	0.83	0.82	0.84	0.83	0.81	0.87
Random Forest	0.85	0.83	0.89	0.79	0.76	0.86

Tabela D.20: Resultados dos modelos criados para a classificação dos quatro níveis de glucose, duas horas após a refeição com a média das componentes do modelo HLS.

## D.5 Modelos Com a Média de LAB

Resultados obtidos na classificação das concentrações e dos níveis de glucose, nos estados de jejum e duas horas após a refeição, baseados na média das três componentes do modelo de cor LAB.

### D.5.1 Concentrações

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.57	0.57	0.65	0.48	0.51	0.48
Support Vector Machines	0.57	0.58	0.60	0.51	0.55	0.52
Random Forest	0.45	0.46	0.46	0.49	0.51	0.50
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.57	0.57	0.59	0.51	0.53	0.52
Support Vector Machines	0.57	0.56	0.64	0.53	0.56	0.53
Random Forest	0.39	0.44	0.36	0.40	0.45	0.37

Tabela D.21: Resultados dos modelos criados para a classificação das concentrações da glucose com a média das componentes do modelo LAB.

### D.5.2 Cinco Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.67	0.68	0.66	0.70	0.71	0.73
Support Vector Machines	0.68	0.68	0.68	0.70	0.70	0.72
Random Forest	0.66	0.66	0.67	0.69	0.70	0.69
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.68	0.68	0.68	0.74	0.74	0.76
Support Vector Machines	0.70	0.69	0.71	0.72	0.72	0.74
Random Forest	0.66	0.67	0.66	0.72	0.73	0.73

Tabela D.22: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com a média das componentes do modelo LAB.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.74	0.73	0.77	0.76	0.76	0.78
Support Vector Machines	0.74	0.74	0.76	0.76	0.75	0.78
Random Forest	0.75	0.74	0.78	0.73	0.72	0.78
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.72	0.69	0.76	0.79	0.78	0.82
Support Vector Machines	0.76	0.74	0.78	0.76	0.76	0.78
Random Forest	0.74	0.74	0.74	0.73	0.74	0.74

Tabela D.23: Resultados dos modelos criados para a classificação dos cinco níveis de glucose, duas horas após a refeição com a média das componentes do modelo LAB.

### D.5.3 Quatro Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.76	0.76	0.76	0.77	0.77	0.79
Support Vector Machines	0.73	0.73	0.74	0.77	0.77	0.78
Random Forest	0.73	0.73	0.73	0.76	0.76	0.77
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.79	0.79	0.79	0.79	0.79	0.79
Support Vector Machines	0.73	0.73	0.73	0.75	0.75	0.77
Random Forest	0.77	0.78	0.77	0.79	0.79	0.80

Tabela D.24: Resultados dos modelos criados para a classificação dos quatro níveis de glucose em jejum com a média das componentes do modelo LAB.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.86	0.86	0.87	0.86	0.85	0.87
Support Vector Machines	0.85	0.84	0.86	0.87	0.85	0.90
Random Forest	0.85	0.83	0.90	0.82	0.79	0.89
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.87	0.86	0.88	0.85	0.84	0.88
Support Vector Machines	0.86	0.84	0.88	0.86	0.84	0.89
Random Forest	0.82	0.83	0.82	0.85	0.84	0.86

Tabela D.25: Resultados dos modelos criados para a classificação dos quatro níveis de glucose, duas horas após a refeição com a média das componentes do modelo LAB.

## D.6 Modelos Com a Média de YUV

Resultados obtidos na classificação das concentrações e dos níveis de glucose, nos estados de jejum e duas horas após a refeição, baseados na média das três componentes do modelo de cor YUV.

### D.6.1 Concentrações

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.57	0.56	0.62	0.48	0.51	0.48
Support Vector Machines	0.50	0.49	0.53	0.50	0.52	0.52
Random Forest	0.45	0.45	0.54	0.46	0.49	0.47
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.57	0.58	0.59	0.52	0.55	0.53
Support Vector Machines	0.56	0.55	0.58	0.52	0.54	0.53
Random Forest	0.42	0.46	0.51	0.44	0.48	0.42

Tabela D.26: Resultados dos modelos criados para a classificação das concentrações da glucose com a média das componentes do modelo YUV.

### D.6.2 Cinco Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.69	0.67	0.73	0.75	0.74	0.78
Support Vector Machines	0.74	0.73	0.75	0.74	0.73	0.77
Random Forest	0.74	0.72	0.78	0.71	0.70	0.76
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.72	0.70	0.76	0.79	0.78	0.82
Support Vector Machines	0.77	0.76	0.80	0.73	0.73	0.76
Random Forest	0.75	0.74	0.76	0.73	0.72	0.76

Tabela D.27: Resultados dos modelos criados para a classificação dos cinco níveis de glucose, duas horas após a refeição com a média das componentes do modelo YUV.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.68	0.69	0.70	0.68	0.70	0.71
Support Vector Machines	0.67	0.68	0.68	0.70	0.71	0.74
Random Forest	0.66	0.67	0.66	0.67	0.69	0.70
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.71	0.71	0.72	0.74	0.75	0.77
Support Vector Machines	0.54	0.54	0.56	0.49	0.51	0.51
Random Forest	0.55	0.54	0.56	0.58	0.58	0.60

Tabela D.28: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com a média das componentes do modelo YUV.

### D.6.3 Quatro Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.75	0.76	0.75	0.78	0.78	0.79
Support Vector Machines	0.68	0.68	0.69	0.72	0.73	0.74
Random Forest	0.68	0.69	0.70	0.70	0.71	0.73
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.76	0.75	0.76	0.79	0.79	0.80
Support Vector Machines	0.74	0.74	0.74	0.73	0.73	0.74
Random Forest	0.78	0.79	0.78	0.75	0.75	0.76

Tabela D.29: Resultados dos modelos criados para a classificação dos quatro níveis de glucose em jejum com a média das componentes do modelo YUV.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.84	0.83	0.86	0.86	0.83	0.89
Support Vector Machines	0.85	0.84	0.87	0.85	0.84	0.89
Random Forest	0.86	0.84	0.89	0.81	0.79	0.87
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.87	0.86	0.88	0.85	0.83	0.89
Support Vector Machines	0.85	0.84	0.87	0.86	0.84	0.89
Random Forest	0.88	0.87	0.89	0.84	0.83	0.86

Tabela D.30: Resultados dos modelos criados para a classificação dos quatro níveis de glucose, duas horas após a refeição com a média das componentes do modelo YUV.

## D.7 Modelos Com a Média, Desvio padrão e Assimetria RGB

Resultados obtidos na classificação das concentrações e dos níveis de glucose, nos estados de jejum e duas horas após a refeição, baseados na média, desvio padrão e assimetria das três componentes do modelo de cor RGB.

### D.7.1 Concentrações

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.38	0.37	0.39	0.39	0.42	0.39
Support Vector Machines	0.33	0.32	0.33	0.40	0.42	0.42
Random Forest	0.31	0.33	0.38	0.35	0.37	0.35
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.53	0.55	0.55	0.50	0.53	0.50
Support Vector Machines	0.46	0.45	0.58	0.36	0.39	0.38
Random Forest	0.38	0.40	0.36	0.34	0.38	0.32

Tabela D.31: Resultados dos modelos criados para a classificação das concentrações da glucose com a média, desvio padrão e assimetria das componentes do modelo RGB.

### D.7.2 Cinco Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.55	0.55	0.58	0.56	0.57	0.60
Support Vector Machines	0.53	0.52	0.56	0.54	0.54	0.59
Random Forest	0.56	0.57	0.59	0.56	0.57	0.61
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.65	0.65	0.66	0.68	0.69	0.70
Support Vector Machines	0.65	0.63	0.67	0.68	0.67	0.72
Random Forest	0.63	0.64	0.64	0.64	0.65	0.68

Tabela D.32: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com a média, desvio padrão e assimetria das componentes do modelo RGB.



## D.7. MODELOS COM A MÉDIA, DESVIO PADRÃO E ASSIMETRIA RGB

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.64	0.63	0.68	0.61	0.59	0.66
Support Vector Machines	0.62	0.59	0.69	0.58	0.56	0.66
Random Forest	0.64	0.63	0.68	0.62	0.61	0.67
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.68	0.68	0.67	0.73	0.73	0.76
Support Vector Machines	0.65	0.62	0.71	0.62	0.59	0.70
Random Forest	0.75	0.75	0.77	0.71	0.70	0.77

Tabela D.33: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com a média, desvio padrão e assimetria das componentes do modelo RGB.

### D.7.3 Quatro Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.65	0.65	0.65	0.64	0.65	0.64
Support Vector Machines	0.62	0.62	0.70	0.61	0.61	0.67
Random Forest	0.62	0.63	0.62	0.64	0.64	0.66
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.76	0.77	0.76	0.77	0.77	0.79
Support Vector Machines	0.76	0.75	0.78	0.77	0.76	0.81
Random Forest	0.73	0.73	0.74	0.73	0.73	0.75

Tabela D.34: Resultados dos modelos criados para a classificação dos quatro níveis de glucose em jejum com a média, desvio padrão e assimetria das componentes do modelo RGB.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.74	0.74	0.75	0.69	0.67	0.79
Support Vector Machines	0.72	0.70	0.76	0.63	0.60	0.73
Random Forest	0.69	0.71	0.68	0.69	0.68	0.74
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.83	0.83	0.83	0.81	0.80	0.85
Support Vector Machines	0.72	0.71	0.74	0.74	0.73	0.79
Random Forest	0.76	0.74	0.77	0.68	0.68	0.73

Tabela D.35: Resultados dos modelos criados para a classificação dos quatro níveis de glucose, duas horas após a refeição com a média, desvio padrão e assimetria das componentes do modelo RGB.

## D.8 Modelos Com a Média, Desvio padrão e Assimetria HSV

Resultados obtidos na classificação das concentrações e dos níveis de glucose, nos estados de jejum e duas horas após a refeição, baseados na média, desvio padrão e assimetria das três componentes do modelo de cor HSV.

### D.8.1 Concentrações

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.41	0.42	0.43	0.39	0.40	0.39
Support Vector Machines	0.41	0.43	0.41	0.38	0.40	0.39
Random Forest	0.32	0.36	0.38	0.31	0.35	0.29
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.62	0.62	0.62	0.59	0.62	0.60
Support Vector Machines	0.53	0.51	0.60	0.59	0.60	0.61
Random Forest	0.48	0.48	0.50	0.40	0.44	0.39

Tabela D.36: Resultados dos modelos criados para a classificação das concentrações da glucose com a média, desvio padrão e assimetria das componentes do modelo HSV.

### D.8.2 Cinco Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.51	0.51	0.51	0.53	0.54	0.550
Support Vector Machines	0.49	0.49	0.49	0.55	0.54	0.58
Random Forest	0.49	0.50	0.50	0.53	0.54	0.56
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.66	0.67	0.67	0.75	0.75	0.78
Support Vector Machines	0.69	0.68	0.71	0.70	0.68	0.75
Random Forest	0.63	0.65	0.68	0.71	0.72	0.78

Tabela D.37: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com a média, desvio padrão e assimetria das componentes do modelo HSV.

## D.8. MODELOS COM A MÉDIA, DESVIO PADRÃO E ASSIMETRIA HSV

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.59	0.58	0.61	0.63	0.62	0.67
Support Vector Machines	0.59	0.56	0.64	0.61	0.58	0.69
Random Forest	0.57	0.55	0.62	0.63	0.62	0.67
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.79	0.79	0.79	0.78	0.78	0.80
Support Vector Machines	0.72	0.70	0.76	0.73	0.71	0.79
Random Forest	0.63	0.61	0.71	0.61	0.62	0.62

Tabela D.38: Resultados dos modelos criados para a classificação dos cinco níveis de glucose, duas horas após a refeição, com a média, desvio padrão e assimetria das componentes do modelo HSV.

### D.8.3 Quatro Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.60	0.60	0.61	0.67	0.67	0.70
Support Vector Machines	0.63	0.63	0.65	0.65	0.64	0.68
Random Forest	0.57	0.57	0.58	0.65	0.65	0.68
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.80	0.80	0.79	0.81	0.82	0.82
Support Vector Machines	0.79	0.78	0.81	0.76	0.75	0.79
Random Forest	0.75	0.74	0.79	0.76	0.76	0.77

Tabela D.39: Resultados dos modelos criados para a classificação dos quatro níveis de glucose em jejum com a média, desvio padrão e assimetria das componentes do modelo HSV.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.80	0.78	0.84	0.76	0.73	0.85
Support Vector Machines	0.75	0.74	0.80	0.71	0.68	0.82
Random Forest	0.78	0.77	0.80	0.77	0.75	0.84
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.84	0.86	0.82	0.87	0.87	0.87
Support Vector Machines	0.83	0.80	0.86	0.76	0.74	0.79
Random Forest	0.85	0.86	0.84	0.82	0.81	0.84

Tabela D.40: Resultados dos modelos criados para a classificação dos quatro níveis de glucose, duas horas após a refeição com a média, desvio padrão e assimetria das componentes do modelo HSV.

## D.9 Modelos Com a Média, Desvio padrão e Assimetria XYZ

Resultados obtidos na classificação das concentrações e dos níveis de glucose, nos estados de jejum e duas horas após a refeição, baseados na média, desvio padrão e assimetria das três componentes do modelo de cor XYZ.

### D.9.1 Concentrações

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.40	0.40	0.41	0.41	0.43	0.41
Support Vector Machines	0.36	0.37	0.37	0.39	0.41	0.39
Random Forest	0.31	0.32	0.31	0.35	0.38	0.35
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.47	0.49	0.47	0.51	0.55	0.52
Support Vector Machines	0.49	0.50	0.50	0.49	0.52	0.50
Random Forest	0.38	0.43	0.36	0.39	0.43	0.39

Tabela D.41: Resultados dos modelos criados para a classificação das concentrações da glucose em all XYZ

### D.9.2 Cinco Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.58	0.58	0.63	0.59	0.61	0.64
Support Vector Machines	0.59	0.60	0.65	0.58	0.58	0.67
Random Forest	0.55	0.56	0.60	0.57	0.58	0.61
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.64	0.64	0.66	0.63	0.64	0.65
Support Vector Machines	0.69	0.69	0.71	0.65	0.64	0.70
Random Forest	0.62	0.63	0.65	0.62	0.63	0.68

Tabela D.42: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum para all XYZ

## D.9. MODELOS COM A MÉDIA, DESVIO PADRÃO E ASSIMETRIA XYZ

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.65	0.63	0.69	0.65	0.64	0.69
Support Vector Machines	0.62	0.59	0.72	0.58	0.56	0.65
Random Forest	0.65	0.63	0.71	0.60	0.58	0.67
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.71	0.69	0.74	0.71	0.69	0.78
Support Vector Machines	0.72	0.69	0.75	0.70	0.68	0.78
Random Forest	0.70	0.67	0.76	0.67	0.65	0.74

Tabela D.43: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum para all XYZ

### D.9.3 Quatro Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.66	0.66	0.66	0.67	0.68	0.69
Support Vector Machines	0.67	0.66	0.72	0.63	0.62	0.69
Random Forest	0.63	0.64	0.63	0.66	0.66	0.68
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.74	0.74	0.74	0.73	0.73	0.75
Support Vector Machines	0.78	0.77	0.79	0.75	0.73	0.78
Random Forest	0.74	0.75	0.74	0.74	0.74	0.75

Tabela D.44: Resultados dos modelos criados para a classificação dos quatro níveis de glucose em jejum para all XYZ

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.75	0.75	0.76	0.70	0.68	0.79
Support Vector Machines	0.73	0.71	0.75	0.63	0.61	0.73
Random Forest	0.75	0.76	0.75	0.71	0.69	0.78
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.81	0.82	0.80	0.76	0.74	0.82
Support Vector Machines	0.69	0.68	0.71	0.71	0.69	0.77
Random Forest	0.81	0.79	0.83	0.75	0.74	0.79

Tabela D.45: Resultados dos modelos criados para a classificação dos quatro níveis de glucose, duas horas após a refeição para all XYZ

## D.10 Modelos Com a Média, Desvio padrão e Assimetria HLS

Resultados obtidos na classificação das concentrações e dos níveis de glucose, nos estados de jejum e duas horas após a refeição, baseados na média, desvio padrão e assimetria das três componentes do modelo de cor HLS.

### D.10.1 Concentrações

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.41	0.41	0.42	0.43	0.45	0.44
Support Vector Machines	0.46	0.45	0.51	0.45	0.48	0.45
Random Forest	0.36	0.39	0.34	0.37	0.42	0.37
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.53	0.54	0.54	0.57	0.59	0.59
Support Vector Machines	0.60	0.60	0.62	0.58	0.59	0.59
Random Forest	0.41	0.43	0.41	0.40	0.45	0.38

Tabela D.46: Resultados dos modelos criados para a classificação das concentrações da glucose com a média, desvio padrão e assimetria das componentes do modelo HLS.

### D.10.2 Cinco Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.63	0.63	0.68	0.65	0.65	0.68
Support Vector Machines	0.62	0.62	0.68	0.65	0.64	0.69
Random Forest	0.61	0.61	0.63	0.64	0.64	0.66
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.72	0.71	0.73	0.76	0.76	0.78
Support Vector Machines	0.74	0.73	0.76	0.72	0.71	0.76
Random Forest	0.51	0.54	0.57	0.58	0.59	0.63

Tabela D.47: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com a média, desvio padrão e assimetria das componentes do modelo HLS.

## D.10. MODELOS COM A MÉDIA, DESVIO PADRÃO E ASSIMETRIA HLS

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.65	0.63	0.68	0.65	0.64	0.67
Support Vector Machines	0.64	0.61	0.71	0.63	0.60	0.71
Random Forest	0.65	0.62	0.71	0.67	0.66	0.71
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.75	0.74	0.77	0.78	0.78	0.79
Support Vector Machines	0.76	0.74	0.78	0.71	0.70	0.75
Random Forest	0.76	0.73	0.80	0.74	0.73	0.79

Tabela D.48: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com a média, desvio padrão e assimetria das componentes do modelo HLS.

### D.10.3 Quatro Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.69	0.69	0.69	0.70	0.70	0.72
Support Vector Machines	0.70	0.71	0.72	0.68	0.68	0.71
Random Forest	0.67	0.67	0.68	0.70	0.70	0.72
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.80	0.80	0.80	0.81	0.81	0.81
Support Vector Machines	0.82	0.82	0.83	0.74	0.73	0.78
Random Forest	0.76	0.77	0.76	0.77	0.77	0.78

Tabela D.49: Resultados dos modelos criados para a classificação dos quatro níveis de glucose em jejum com a média, desvio padrão e assimetria das componentes do modelo HLS.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.84	0.83	0.86	0.78	0.76	0.84
Support Vector Machines	0.84	0.82	0.89	0.77	0.74	0.85
Random Forest	0.84	0.82	0.87	0.78	0.75	0.86
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.83	0.84	0.82	0.86	0.86	0.87
Support Vector Machines	0.81	0.81	0.81	0.77	0.75	0.82
Random Forest	0.54	0.55	0.75	0.51	0.54	0.60

Tabela D.50: Resultados dos modelos criados para a classificação dos quatro níveis de glucose, duas horas após a refeição com a média, desvio padrão e assimetria das componentes do modelo HLS.

## D.11 Modelos Com a Média, Desvio padrão e Assimetria LAB

Resultados obtidos na classificação das concentrações e dos níveis de glucose, nos estados de jejum e duas horas após a refeição, baseados na média, desvio padrão e assimetria das três componentes do modelo de cor LAB.

### D.11.1 Concentrações

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.37	0.37	0.39	0.41	0.43	0.42
Support Vector Machines	0.37	0.37	0.39	0.37	0.39	0.38
Random Forest	0.29	0.33	0.26	0.38	0.41	0.38
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.62	0.63	0.62	0.62	0.65	0.63
Support Vector Machines	0.67	0.67	0.69	0.59	0.60	0.62
Random Forest	0.46	0.49	0.47	0.49	0.51	0.50

Tabela D.51: Resultados dos modelos criados para a classificação das concentrações da glucose com a média, desvio padrão e assimetria das componentes do modelo LAB.

### D.11.2 Cinco Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.56	0.57	0.58	0.56	0.57	0.57
Support Vector Machines	0.52	0.53	0.52	0.53	0.53	0.56
Random Forest	0.50	0.53	0.48	0.53	0.54	0.54
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.73	0.73	0.75	0.77	0.77	0.79
Support Vector Machines	0.68	0.66	0.74	0.64	0.62	0.72
Random Forest	0.72	0.72	0.75	0.74	0.74	0.77

Tabela D.52: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com a média, desvio padrão e assimetria das componentes do modelo LAB.



### D.11. MODELOS COM A MÉDIA, DESVIO PADRÃO E ASSIMETRIA LAB

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.65	0.63	0.68	0.64	0.63	0.67
Support Vector Machines	0.54	0.55	0.59	0.59	0.60	0.66
Random Forest	0.59	0.61	0.59	0.62	0.63	0.66
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.79	0.79	0.80	0.79	0.78	0.82
Support Vector Machines	0.75	0.73	0.78	0.74	0.73	0.79
Random Forest	0.78	0.78	0.79	0.77	0.77	0.78

Tabela D.53: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com a média, desvio padrão e assimetria das componentes do modelo LAB.

#### D.11.3 Quatro Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.68	0.68	0.68	0.68	0.68	0.69
Support Vector Machines	0.70	0.69	0.71	0.67	0.67	0.73
Random Forest	0.71	0.70	0.72	0.68	0.68	0.71
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.83	0.83	0.83	0.82	0.82	0.82
Support Vector Machines	0.79	0.79	0.79	0.76	0.75	0.77
Random Forest	0.81	0.81	0.82	0.80	0.80	0.82

Tabela D.54: Resultados dos modelos criados para a classificação dos quatro níveis de glucose em jejum com a média, desvio padrão e assimetria das componentes do modelo LAB.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.79	0.78	0.81	0.75	0.73	0.82
Support Vector Machines	0.76	0.77	0.77	0.70	0.68	0.82
Random Forest	0.79	0.78	0.80	0.76	0.74	0.85
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.87	0.89	0.86	0.83	0.83	0.84
Support Vector Machines	0.79	0.77	0.82	0.73	0.71	0.79
Random Forest	0.82	0.84	0.81	0.84	0.84	0.85

Tabela D.55: Resultados dos modelos criados para a classificação dos quatro níveis de glucose, duas horas após a refeição com a média, desvio padrão e assimetria das componentes do modelo LAB.

## D.12 Modelos Com a Média, Desvio padrão e Assimetria YUV

Resultados obtidos na classificação das concentrações e dos níveis de glucose, nos estados de jejum e duas horas após a refeição, baseados na média, desvio padrão e assimetria das três componentes do modelo de cor YUV.

### D.12.1 Concentrações

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.37	0.37	0.37	0.41	0.42	0.42
Support Vector Machines	0.38	0.39	0.41	0.38	0.41	0.41
Random Forest	0.33	0.35	0.40	0.38	0.41	0.38
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.65	0.66	0.65	0.64	0.66	0.66
Support Vector Machines	0.71	0.72	0.73	0.69	0.69	0.73
Random Forest	0.42	0.43	0.47	0.40	0.43	0.40

Tabela D.56: Resultados dos modelos criados para a classificação das concentrações da glucose com a média, desvio padrão e assimetria das componentes do modelo YUV.

### D.12.2 Cinco Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.55	0.54	0.58	0.55	0.56	0.57
Support Vector Machines	0.54	0.54	0.56	0.49	0.51	0.51
Random Forest	0.55	0.54	0.56	0.58	0.58	0.60
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.77	0.76	0.78	0.79	0.79	0.80
Support Vector Machines	0.73	0.72	0.75	0.75	0.74	0.78
Random Forest	0.72	0.73	0.75	0.76	0.76	0.80

Tabela D.57: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com a média, desvio padrão e assimetria das componentes do modelo YUV.

## D.12. MODELOS COM A MÉDIA, DESVIO PADRÃO E ASSIMETRIA YUV

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.55	0.54	0.58	0.58	0.57	0.63
Support Vector Machines	0.53	0.53	0.59	0.54	0.54	0.64
Random Forest	0.55	0.57	0.57	0.57	0.58	0.60
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.81	0.79	0.84	0.79	0.78	0.82
Support Vector Machines	0.77	0.74	0.82	0.75	0.73	0.79
Random Forest	0.79	0.78	0.82	0.75	0.74	0.81

Tabela D.58: Resultados dos modelos criados para a classificação dos cinco níveis de glucose após duas horas com a média, desvio padrão e assimetria das componentes do modelo YUV.

## D.12.3 Quatro Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.64	0.63	0.65	0.61	0.61	0.64
Support Vector Machines	0.65	0.65	0.73	0.59	0.60	0.70
Random Forest	0.64	0.63	0.66	0.63	0.63	0.64
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.84	0.85	0.84	0.84	0.84	0.85
Support Vector Machines	0.84	0.84	0.85	0.81	0.79	0.84
Random Forest	0.80	0.80	0.80	0.80	0.81	0.82

Tabela D.59: Resultados dos modelos criados para a classificação dos quatro níveis de glucose em jejum com a média, desvio padrão e assimetria das componentes do modelo YUV.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.74	0.74	0.75	0.66	0.66	0.72
Support Vector Machines	0.72	0.71	0.78	0.62	0.62	0.78
Random Forest	0.73	0.75	0.76	0.66	0.65	0.74
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.87	0.87	0.86	0.84	0.84	0.85
Support Vector Machines	0.85	0.84	0.87	0.74	0.73	0.78
Random Forest	0.86	0.88	0.85	0.86	0.86	0.86

Tabela D.60: Resultados dos modelos criados para a classificação dos quatro níveis de glucose, duas horas após a refeição com a média, desvio padrão e assimetria das componentes do modelo YUV.

## D.13 Rácio R/G

Resultados obtidos na classificação das concentrações e dos níveis de glucose, nos estados de jejum e duas horas após a refeição, baseados no rácio entre a média das componentes vermelha e verde do modelo de cor RGB, sem a aplicação de PCA.

### D.13.1 Concentrações

Classificador	Conjunto de Teste			Cross-Validation		
	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.38	0.39	0.37	0.37	0.39	0.37
Support Vector Machines	0.40	0.43	0.39	0.38	0.42	0.38
Random Forest	0.35	0.40	0.32	0.35	0.40	0.33

Tabela D.61: Resultados dos modelos criados para a classificação das concentrações da glucose em RG Ratio sem PCA

### D.13.2 Cinco Níveis

Classificador	Conjunto de Teste			Cross-Validation		
	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.56	0.57	0.56	0.61	0.62	0.62
Support Vector Machines	0.57	0.60	0.58	0.59	0.61	0.62
Random Forest	0.55	0.59	0.52	0.60	0.61	0.62

Tabela D.62: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com o rácio entre a média das componentes vermelha e verde do modelo RGB.

Classificador	Conjunto de Teste			Cross-Validation		
	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.67	0.67	0.70	0.69	0.69	0.74
Support Vector Machines	0.66	0.66	0.73	0.67	0.67	0.74
Random Forest	0.65	0.67	0.65	0.67	0.68	0.70

Tabela D.63: Resultados dos modelos criados para a classificação dos cinco níveis de glucose após duas horas com o rácio entre a média das componentes vermelha e verde do modelo RGB.

#### D.14. MÉDIA DOS COMPONENTES V E S DO MODELO HSV

Classificador	Conjunto de Teste			Cross-Validation		
	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.76	0.76	0.77	0.74	0.74	0.77
Support Vector Machines	0.76	0.76	0.81	0.73	0.74	0.80
Random Forest	0.76	0.75	0.80	0.73	0.74	0.80

Tabela D.64: Resultados dos modelos criados para a classificação dos quatro níveis de glucose em jejum com o rácio entre a média das componentes vermelha e verde do modelo RGB.

#### D.13.3 Quatro Níveis

Classificador	Conjunto de Teste			Cross-Validation		
	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.85	0.86	0.84	0.83	0.83	0.85
Support Vector Machines	0.84	0.85	0.84	0.84	0.83	0.86
Random Forest	0.84	0.86	0.83	0.84	0.83	0.87

Tabela D.65: Resultados dos modelos criados para a classificação dos quatro níveis de glucose após duas horas com o rácio entre a média das componentes vermelha e verde do modelo RGB.

### D.14 Média dos Componentes V e S do Modelo HSV

Resultados obtidos na classificação das concentrações e dos níveis de glucose, nos estados de jejum e duas horas após a refeição, baseados na média dos componentes V e S do modelo HSV.

#### D.14.1 Concentrações

Classificador	Conjunto de Teste			Cross-Validation		
	F1-Score	Recall	Precision	F1-Score	Recall	Precision
Com PCA						
K-Nearest Neighbours	0.53	0.52	0.54	0.49	0.50	0.50
Support Vector Machines	0.52	0.53	0.59	0.48	0.50	0.49
Random Forest	0.46	0.49	0.46	0.41	0.45	0.42
Sem PCA						
K-Nearest Neighbours	0.53	0.52	0.54	0.49	0.50	0.50
Support Vector Machines	0.52	0.53	0.59	0.48	0.50	0.49
Random Forest	0.31	0.38	0.28	0.35	0.41	0.34

Tabela D.66: Resultados dos modelos criados para a classificação das concentrações da glucose com a média dos componentes V e S do modelo HSV.

**D.14.2 Cinco Níveis**

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.64	0.64	0.65	0.66	0.67	0.68
Support Vector Machines	0.64	0.64	0.65	0.68	0.68	0.70
Random Forest	0.66	0.66	0.67	0.66	0.66	0.67
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.64	0.64	0.65	0.66	0.67	0.68
Support Vector Machines	0.64	0.64	0.65	0.68	0.68	0.70
Random Forest	0.62	0.63	0.64	0.65	0.65	0.67

Tabela D.67: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com a média dos componentes V e S do modelo HSV.

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.66	0.65	0.69	0.73	0.72	0.75
Support Vector Machines	0.66	0.66	0.72	0.70	0.71	0.77
Random Forest	0.71	0.69	0.77	0.71	0.70	0.77
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.66	0.65	0.69	0.73	0.72	0.75
Support Vector Machines	0.66	0.66	0.72	0.70	0.71	0.77
Random Forest	0.72	0.71	0.74	0.66	0.66	0.68

Tabela D.68: Resultados dos modelos criados para a classificação dos cinco níveis de glucose após duas horas com a média dos componentes V e S do modelo HSV.

**D.14.3 Quatro Níveis**

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.77	0.77	0.77	0.73	0.74	0.75
Support Vector Machines	0.71	0.71	0.71	0.73	0.74	0.74
Random Forest	0.74	0.73	0.75	0.74	0.74	0.75
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.77	0.77	0.77	0.73	0.74	0.75
Support Vector Machines	0.71	0.71	0.71	0.73	0.74	0.74
Random Forest	0.74	0.74	0.74	0.75	0.75	0.76

Tabela D.69: Resultados dos modelos criados para a classificação dos quatro níveis de glucose em jejum com a média dos componentes V e S do modelo HSV.

## D.15. MÉDIA DOS COMPONENTES L E S DO MODELO HLS

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.85	0.85	0.86	0.87	0.85	0.90
Support Vector Machines	0.85	0.85	0.85	0.86	0.85	0.87
Random Forest	0.87	0.85	0.90	0.82	0.80	0.90
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.85	0.85	0.86	0.87	0.85	0.90
Support Vector Machines	0.85	0.85	0.85	0.86	0.85	0.87
Random Forest	0.84	0.86	0.83	0.83	0.82	0.84

Tabela D.70: Resultados dos modelos criados para a classificação dos quatro níveis de glucose após duas horas com a média dos componentes V e S do modelo HSV.

## D.15 Média dos Componentes L e S do Modelo HLS

Resultados obtidos na classificação das concentrações e dos níveis de glucose, nos estados de jejum e duas horas após a refeição, baseados na média dos componentes L e S do modelo HLS.

### D.15.1 Concentrações

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.52	0.51	0.55	0.45	0.47	0.46
Support Vector Machines	0.52	0.51	0.57	0.43	0.45	0.44
Random Forest	0.43	0.46	0.43	0.40	0.43	0.39
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.52	0.51	0.55	0.45	0.47	0.46
Support Vector Machines	0.52	0.51	0.57	0.43	0.45	0.44
Random Forest	0.44	0.46	0.44	0.43	0.46	0.42

Tabela D.71: Resultados dos modelos criados para a classificação das concentrações da glucose com a média dos componentes L e S do modelo HLS.

## APÊNDICE D. RESULTADOS DA CLASSIFICAÇÃO DA GLUCOSE

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.67	0.67	0.67	0.66	0.66	0.68
Support Vector Machines	0.66	0.65	0.68	0.68	0.68	0.69
Random Forest	0.70	0.70	0.73	0.65	0.66	0.65
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.67	0.67	0.67	0.66	0.66	0.68
Support Vector Machines	0.66	0.65	0.68	0.68	0.68	0.69
Random Forest	0.64	0.64	0.66	0.60	0.60	0.62

Tabela D.72: Resultados dos modelos criados para a classificação dos cinco níveis de glucose em jejum com a média dos componentes L e S do modelo HLS.

### D.15.2 Cinco Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.69	0.67	0.72	0.71	0.72	0.73
Support Vector Machines	0.66	0.64	0.69	0.70	0.69	0.73
Random Forest	0.67	0.68	0.66	0.69	0.69	0.72
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.69	0.67	0.72	0.71	0.72	0.73
Support Vector Machines	0.66	0.64	0.69	0.70	0.69	0.73
Random Forest	0.66	0.64	0.70	0.67	0.67	0.70

Tabela D.73: Resultados dos modelos criados para a classificação dos cinco níveis de glucose após duas horas com a média dos componentes L e S do modelo HLS.

### D.15.3 Quatro Níveis

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.71	0.71	0.71	0.73	0.74	0.74
Support Vector Machines	0.69	0.69	0.70	0.73	0.73	0.75
Random Forest	0.76	0.77	0.76	0.74	0.75	0.75
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.71	0.71	0.71	0.73	0.74	0.74
Support Vector Machines		0.69	0.69	0.70	0.73	0.73
Random Forest	0.73	0.74	0.73	0.67	0.68	0.68

Tabela D.74: Resultados dos modelos criados para a classificação dos quatro níveis de glucose em jejum com a média dos componentes L e S do modelo HLS.



## D.15. MÉDIA DOS COMPONENTES L E S DO MODELO HLS

Classificador	Conjunto de Teste			Cross-Validation		
Com PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.86	0.86	0.87	0.86	0.85	0.89
Support Vector Machines	0.85	0.84	0.86	0.86	0.85	0.89
Random Forest	0.85	0.86	0.84	0.86	0.86	0.89
Sem PCA	F1-Score	Recall	Precision	F1-Score	Recall	Precision
K-Nearest Neighbours	0.86	0.86	0.87	0.86	0.85	0.89
Support Vector Machines	0.85	0.84	0.86	0.86	0.85	0.89
Random Forest	0.86	0.85	0.88	0.81	0.79	0.88

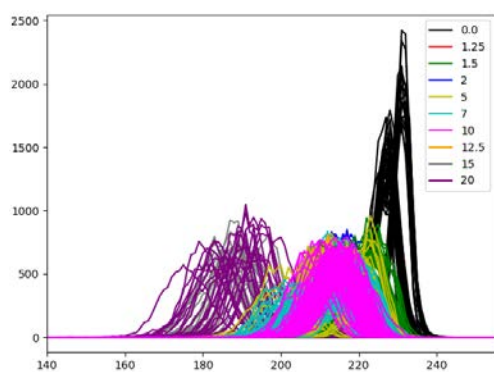
Tabela D.75: Resultados dos modelos criados para a classificação dos quatro níveis de glucose após duas horas com a média dos componentes L e S do modelo HLS.



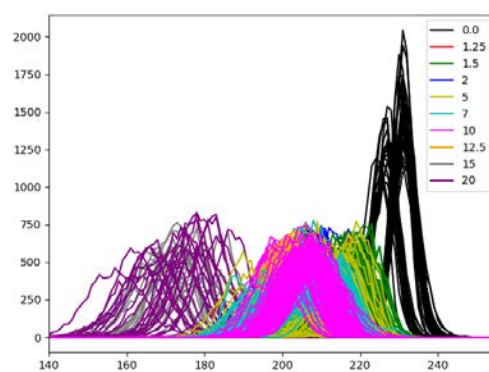
## HISTOGRAMAS

Neste apêndice estão disponíveis os histogramas de cor para os diferentes modelos de cor referentes aos dois conjuntos de dados: glucose e ácido úrico. Cada cor diferente no gráfico corresponde a uma concentração distinta da substância. Estes histogramas foram criados para a observação e identificação de potenciais *features*, para utilização na classificação das substâncias descrita na secção 4.1.1.2.

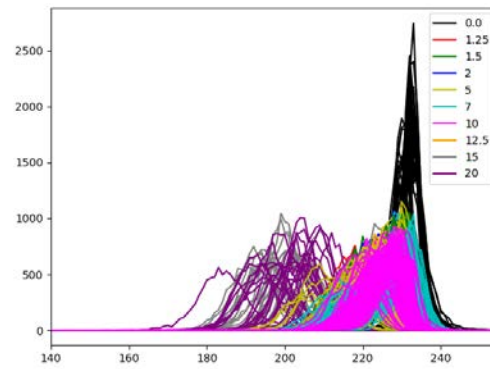
### E.1 Glucose



(a) Componente B

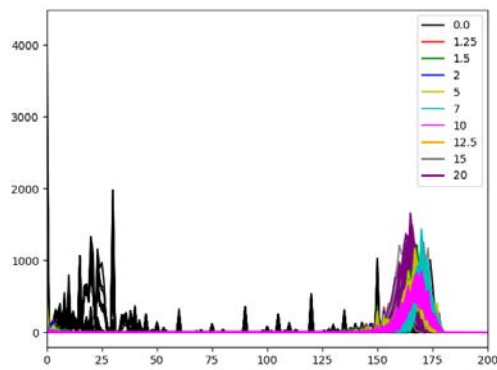


(b) Componente G

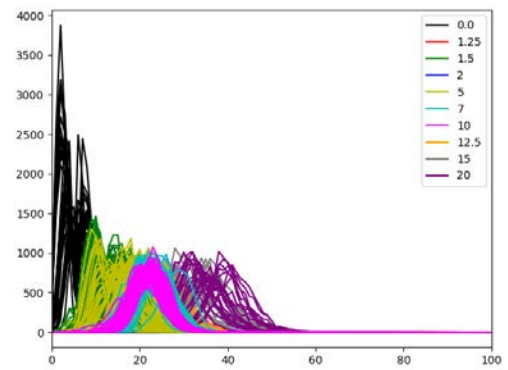


(c) Componente R

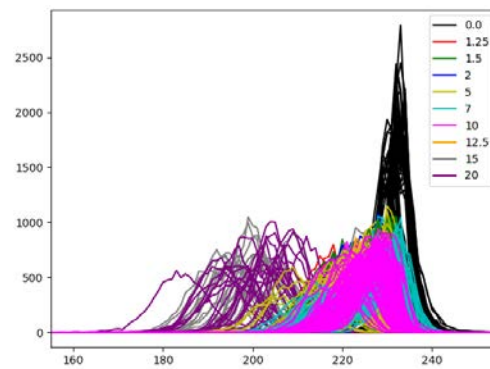
Figura E.1: Histogramas para os três componentes RGB do conjunto de dados da glucose.



(a) Componente H

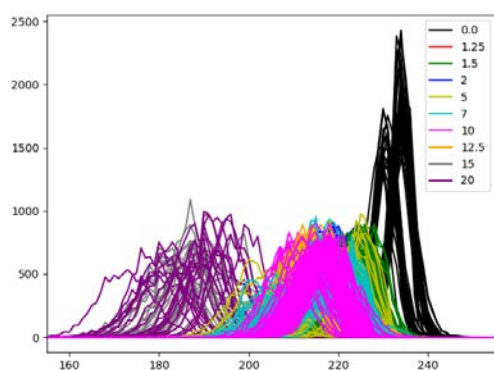


(b) Componente S

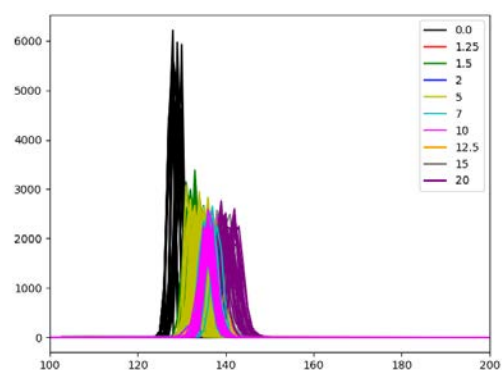


(c) Componente V

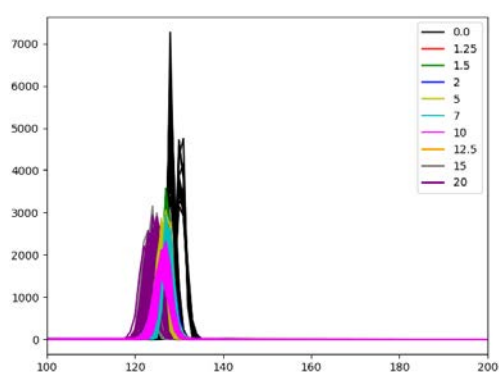
Figura E.2: Histogramas para os três componentes HSV do conjunto de dados da glucose.



(a) Componente L

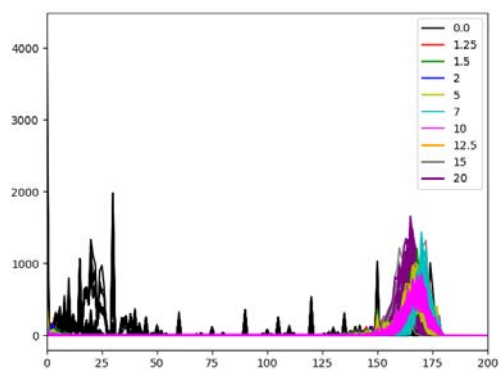


(b) Componente a

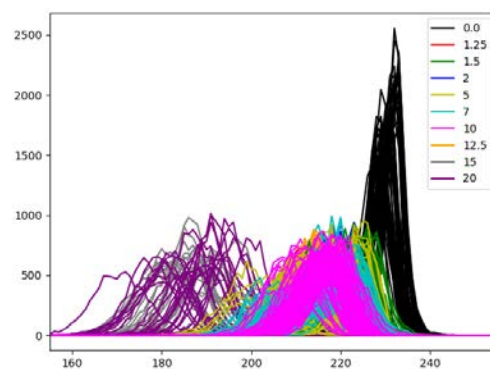


(c) Componente b

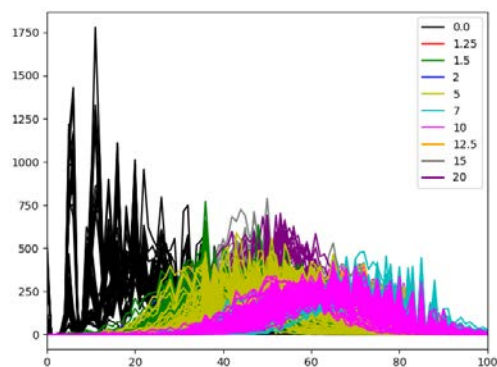
Figura E.3: Histogramas para os três componentes LAB do conjunto de dados da glucose.



(a) Componente H



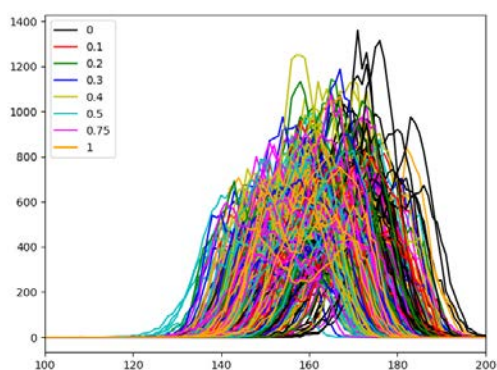
(b) Componente L



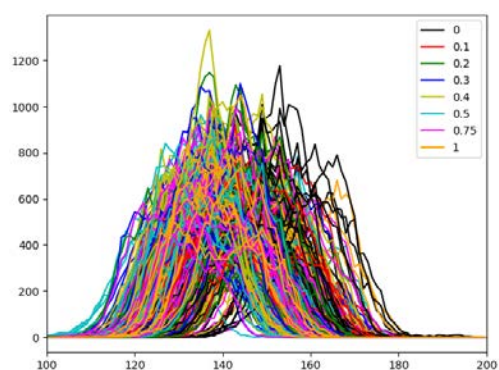
(c) Componente S

Figura E.4: Histogramas para os três componentes HLS do conjunto de dados da glicose.

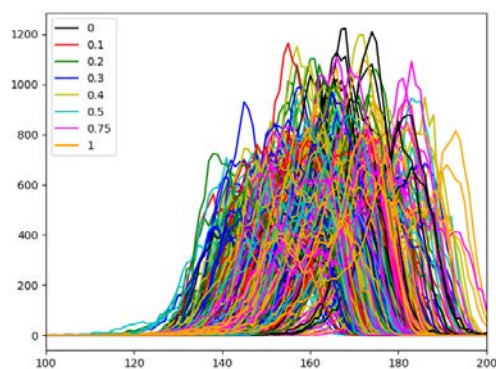
## E.2 Ácido Úrico



(a) Componente B



(b) Componente G



(c) Componente R

Figura E.5: Histogramas para os três componentes RGB do conjunto de dados do ácido úrico.

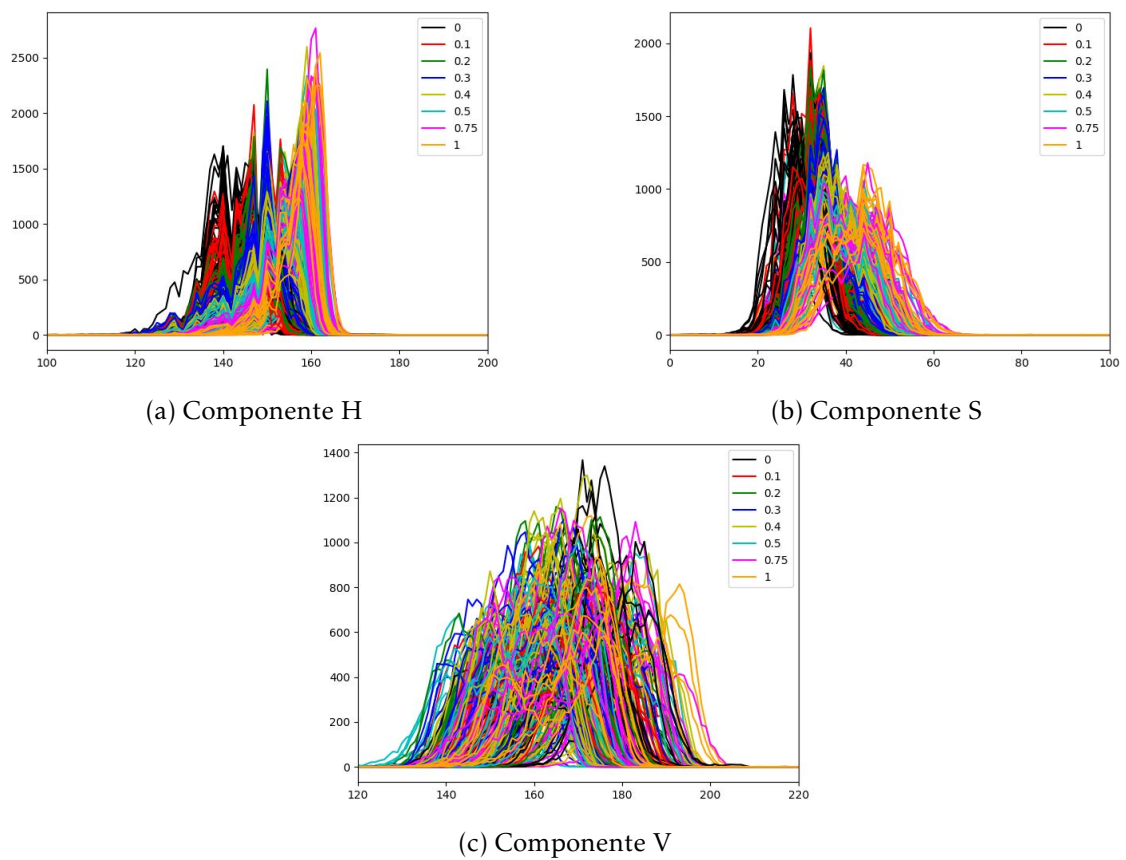
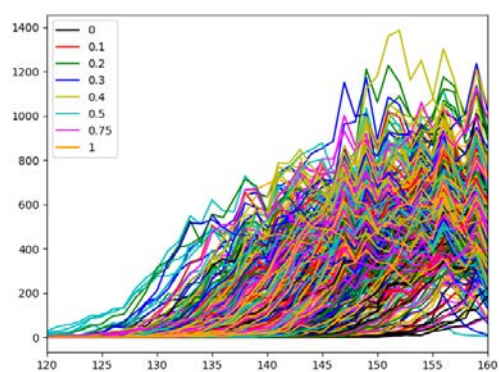
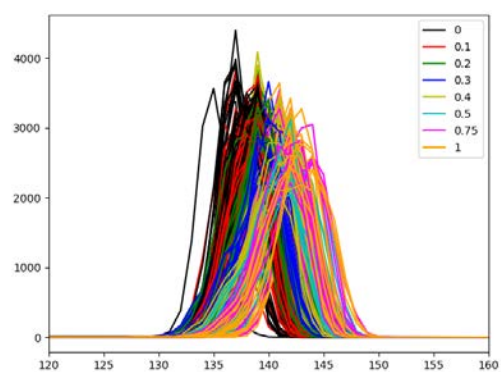


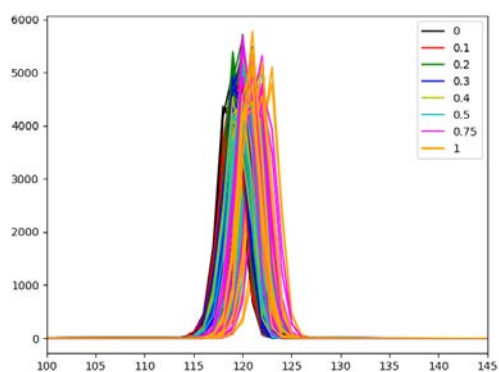
Figura E.6: Histogramas para os três componentes HSV do conjunto de dados do ácido úrico.



(a) Componente L



(b) Componente a



(c) Componente b

Figura E.7: Histogramas para os três componentes Lab do conjunto de dados do ácido úrico.



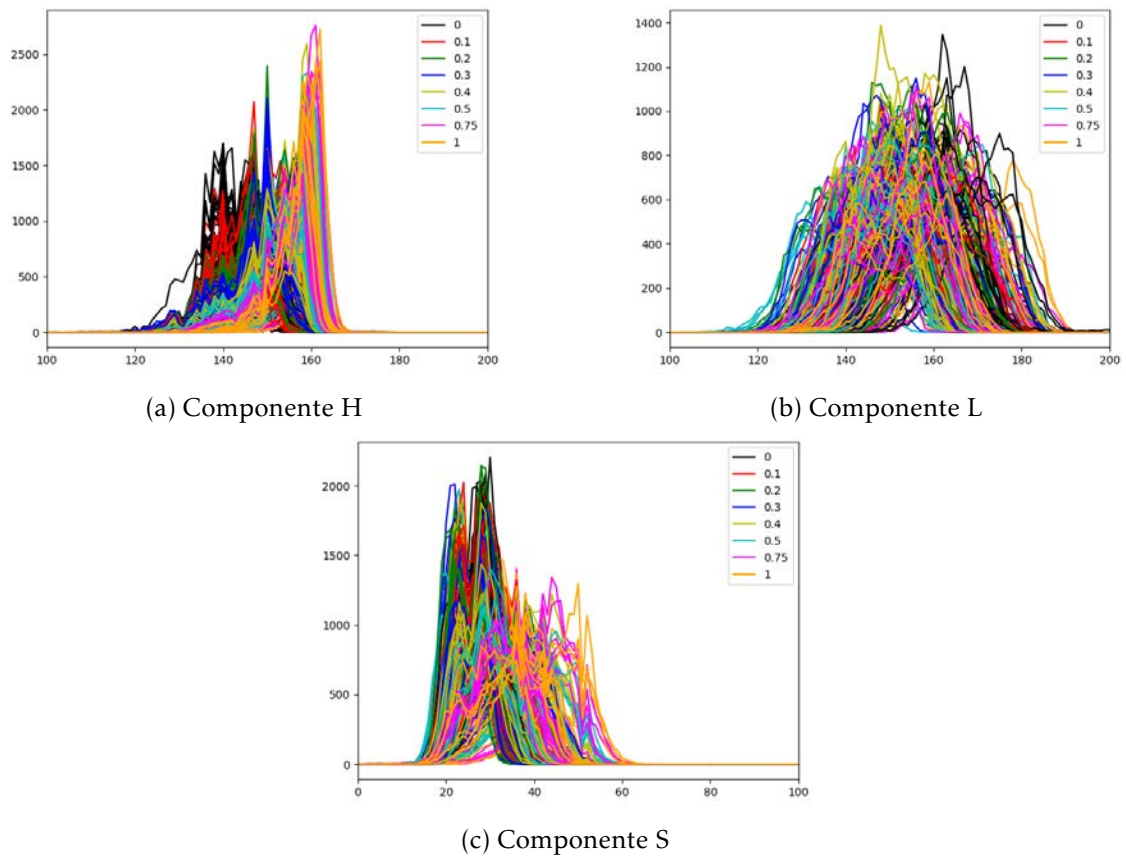


Figura E.8: Histogramas para os três componentes HLS do conjunto de dados do ácido úrico.

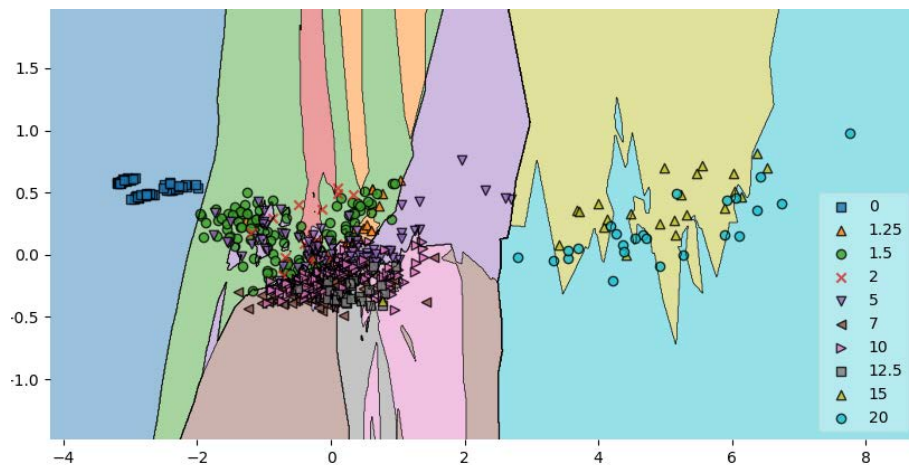




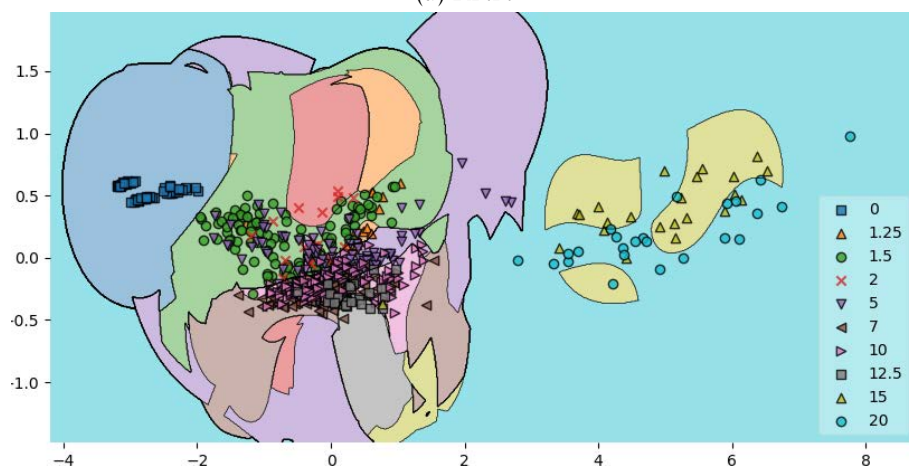
## REGIÕES DE DECISÃO DOS MODELOS DESENVOLVIDOS

Neste apêndice são apresentadas as regiões de decisão obtidas para os modelos desenvolvidos para a classificação da glucose e do ácido úrico, através do processo detalhado na secção 4.1. São apresentados apenas os modelos que permitem o desenho da região, isto é, aqueles que possuem duas *features* ou sofreram redução para duas dimensões através do PCA.

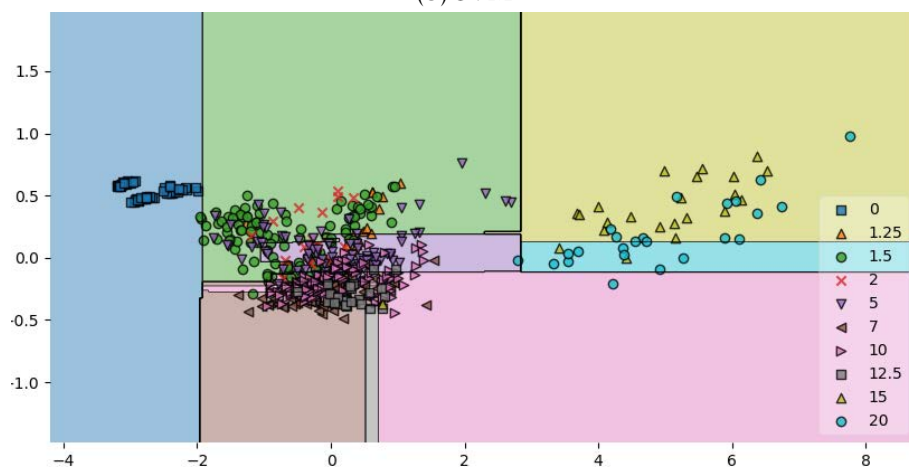
## F.1 Concentrações de Glucose



(a) KNN

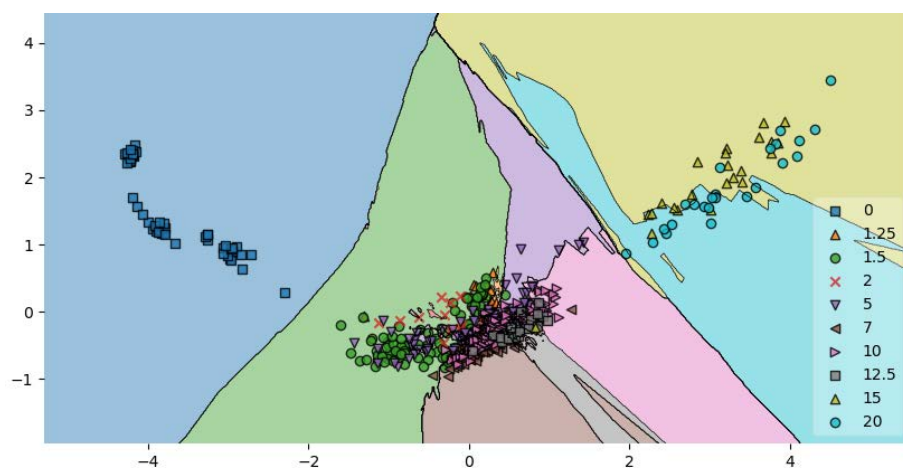


(b) SVM

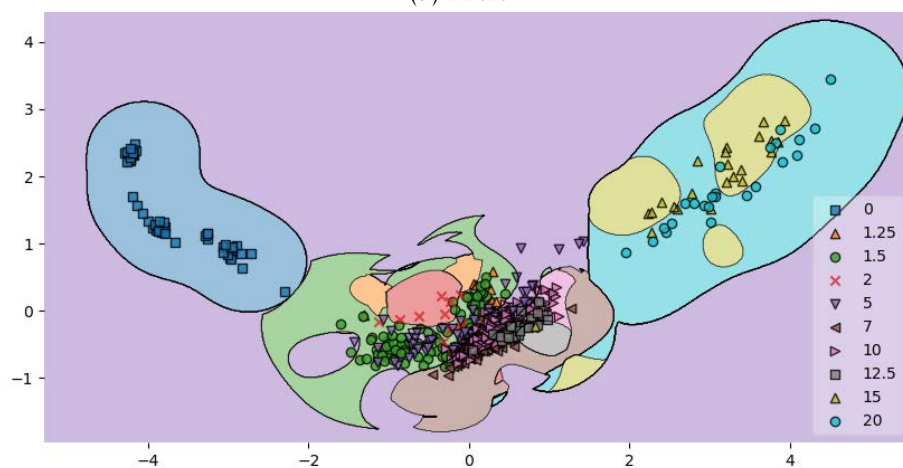


(c) Random Forest

Figura F.1: Regiões de decisão para as concentrações de glucose (mM) dos três algoritmos estudados, baseados na média dos componentes RGB.



(a) KNN



(b) SVM

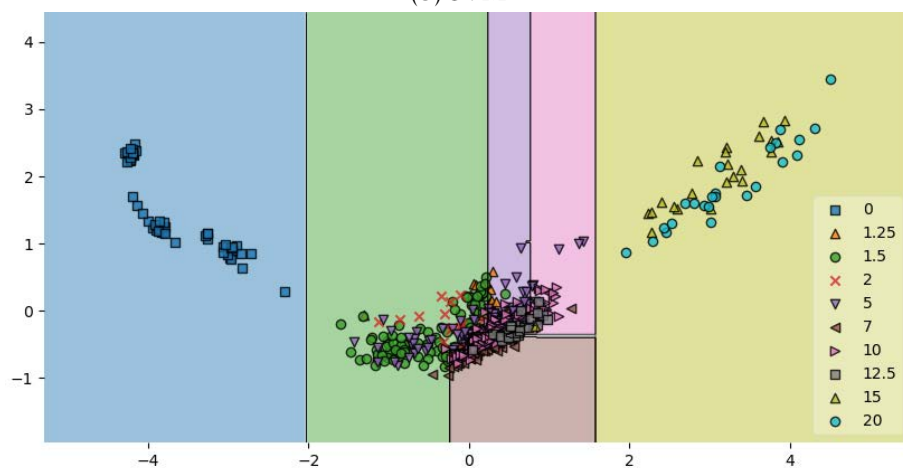
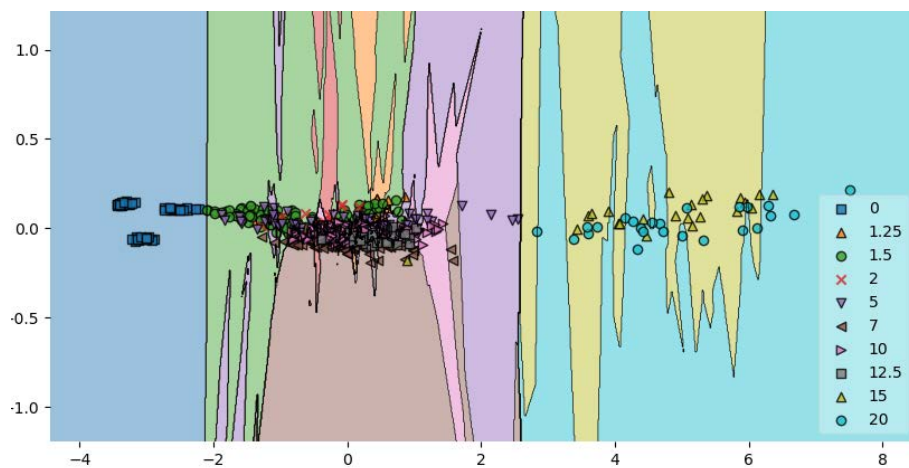
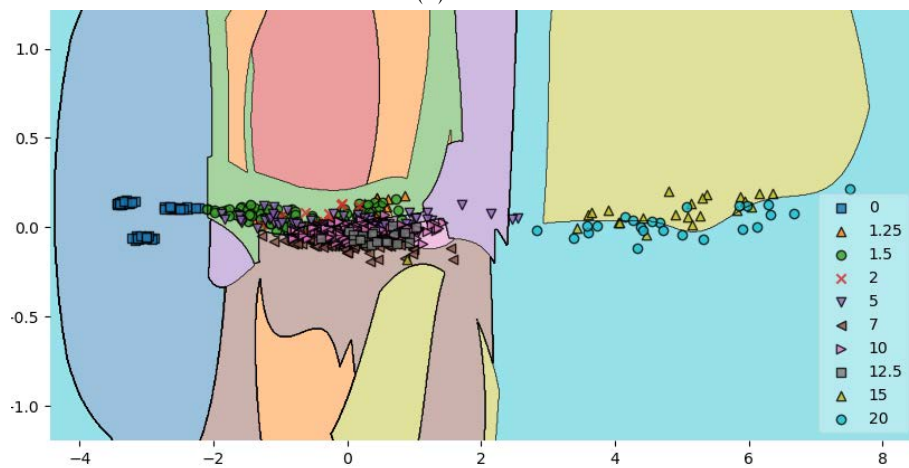
(c) *Random Forest*

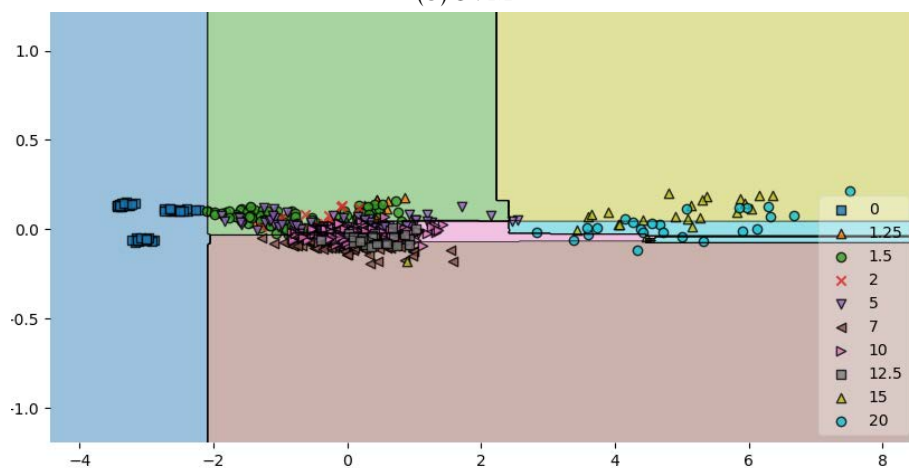
Figura F.2: Regiões de decisão para as concentrações de glucose (mM) dos três algoritmos estudados, baseados na média dos componentes HSV.



(a) KNN

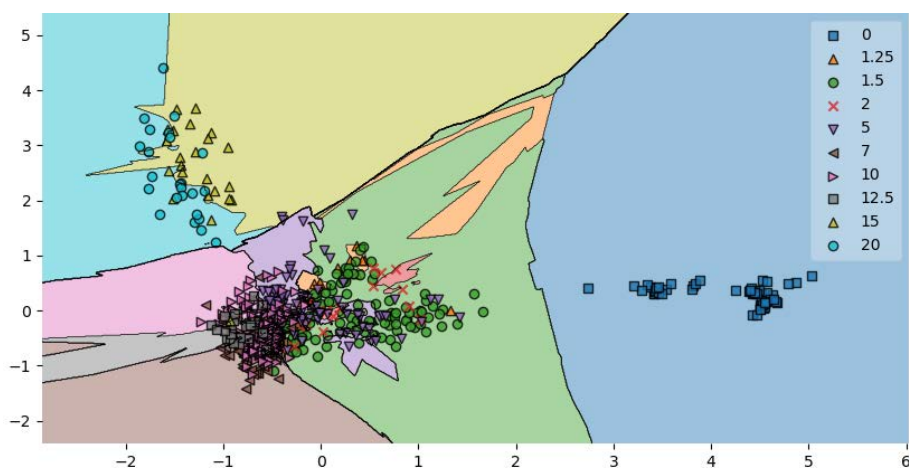


(b) SVM

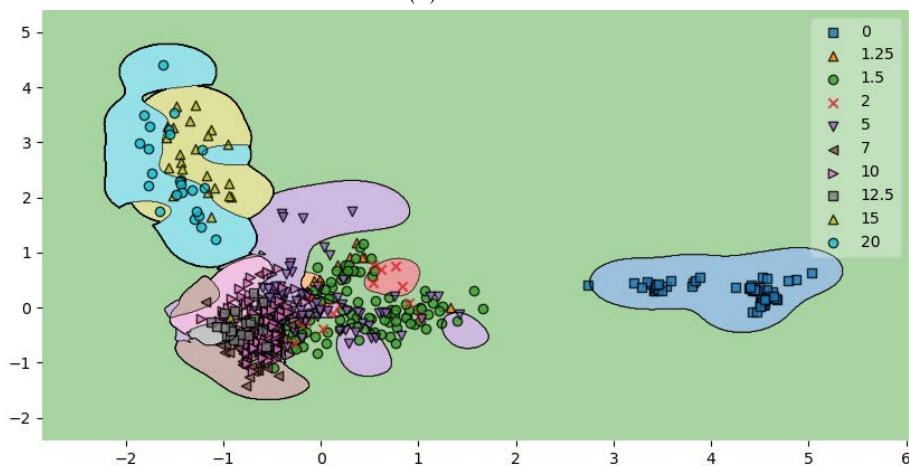


(c) *Random Forest*

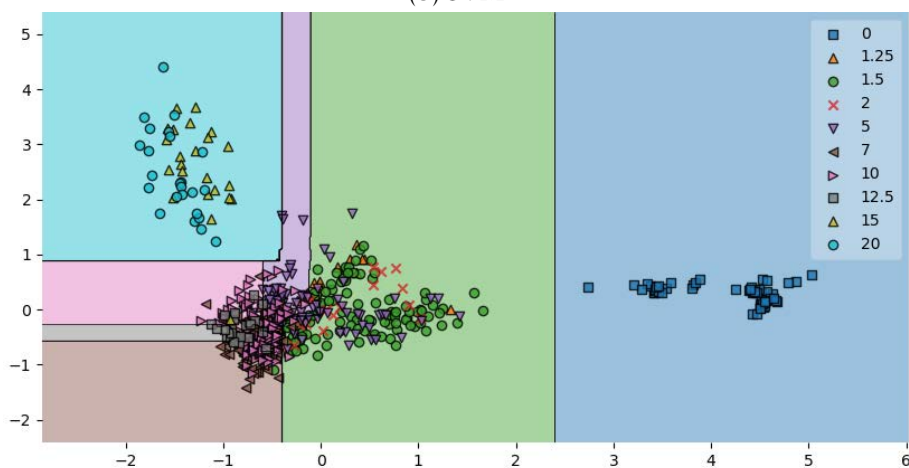
Figura F.3: Regiões de decisão para as concentrações de glucose (mM) dos três algoritmos estudados, baseados na média dos componentes XYZ.



(a) KNN



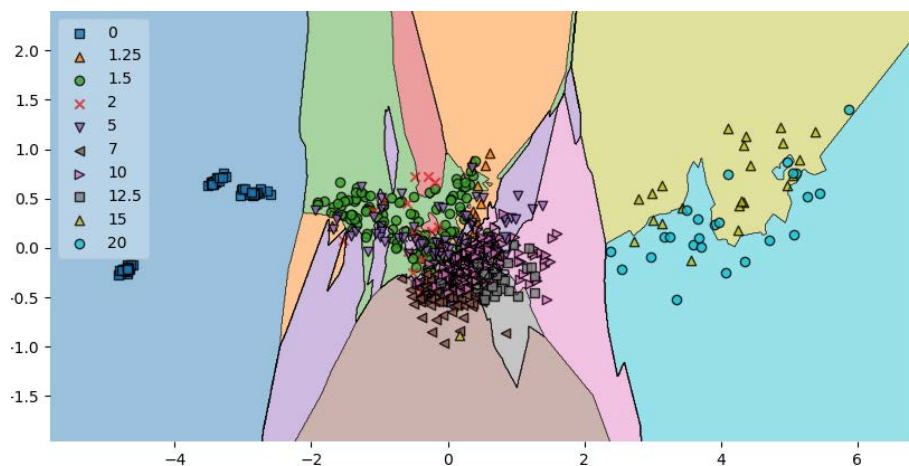
(b) SVM



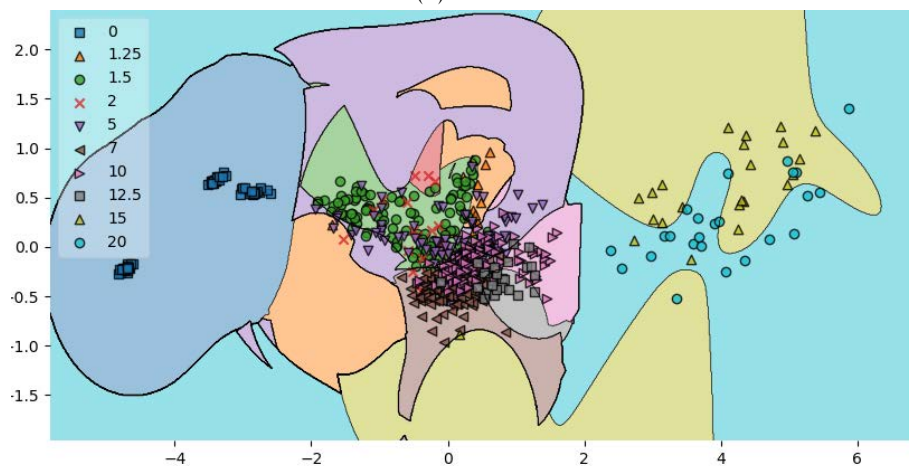
(c) Random Forest

Figura F.4: Regiões de decisão para as concentrações de glucose (mM) dos três algoritmos estudados, baseados na média dos componentes HLS.

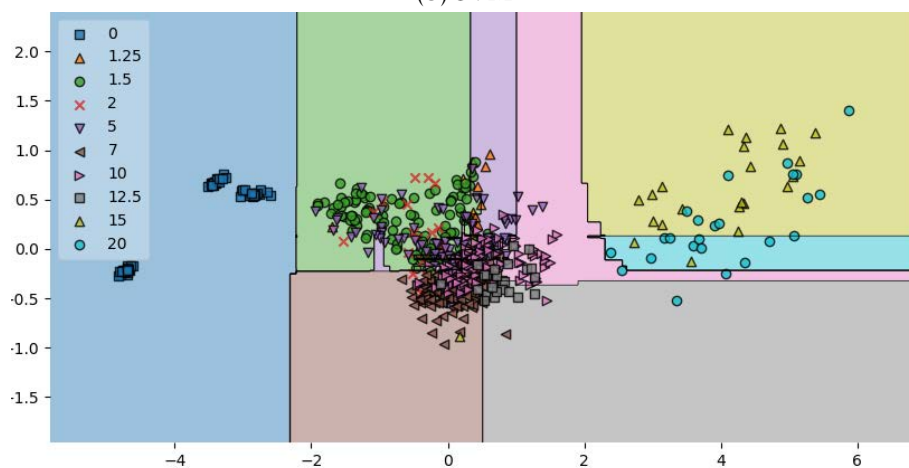




(a) KNN



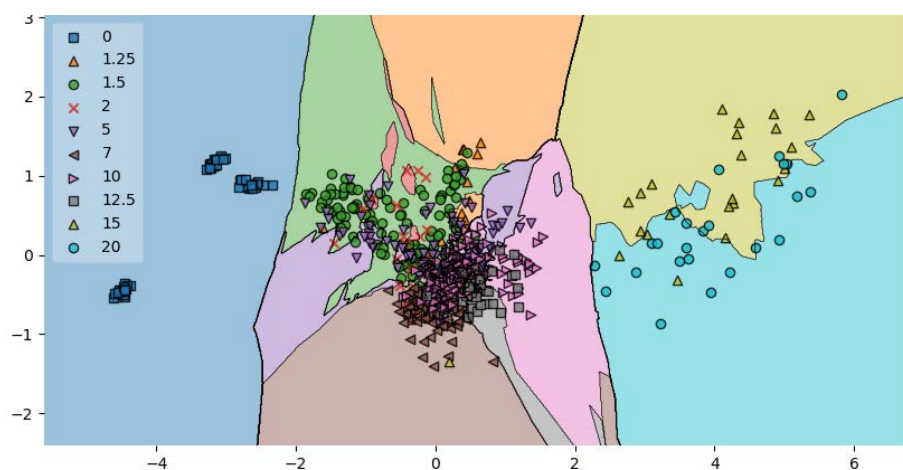
(b) SVM



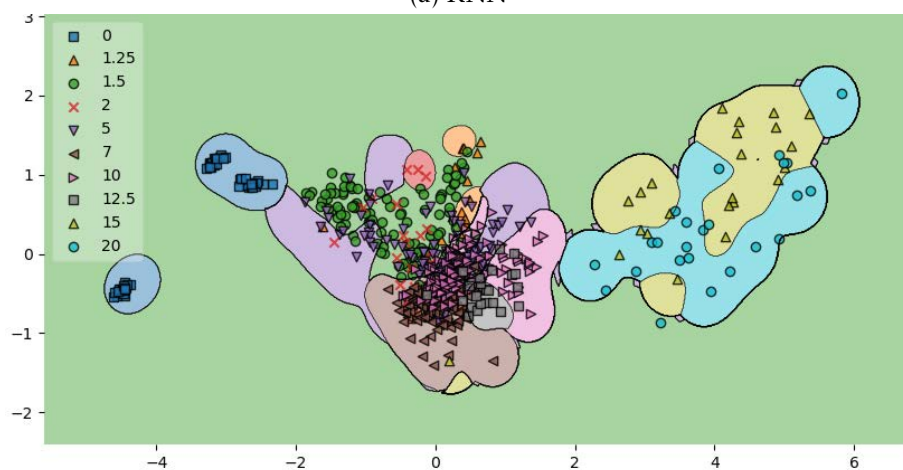
(c) Random Forest

Figura F.5: Regiões de decisão para as concentrações de glucose (mM) dos três algoritmos estudados, baseados na média dos componentes LAB.





(a) KNN



(b) SVM

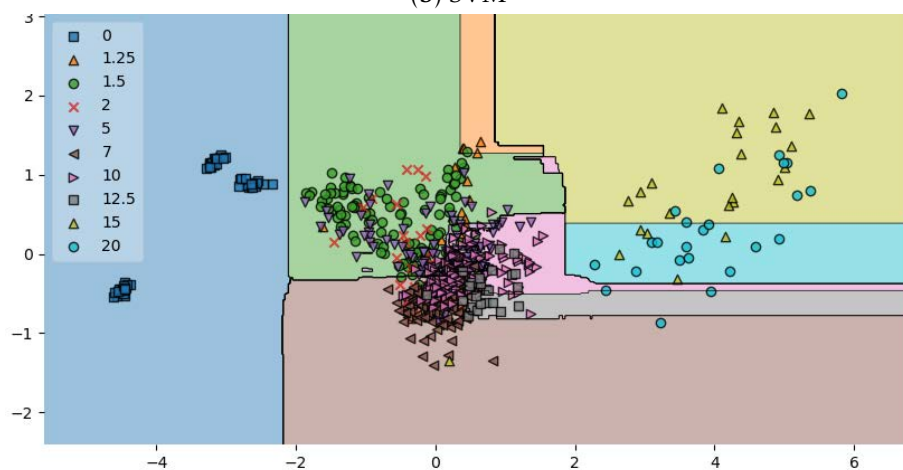
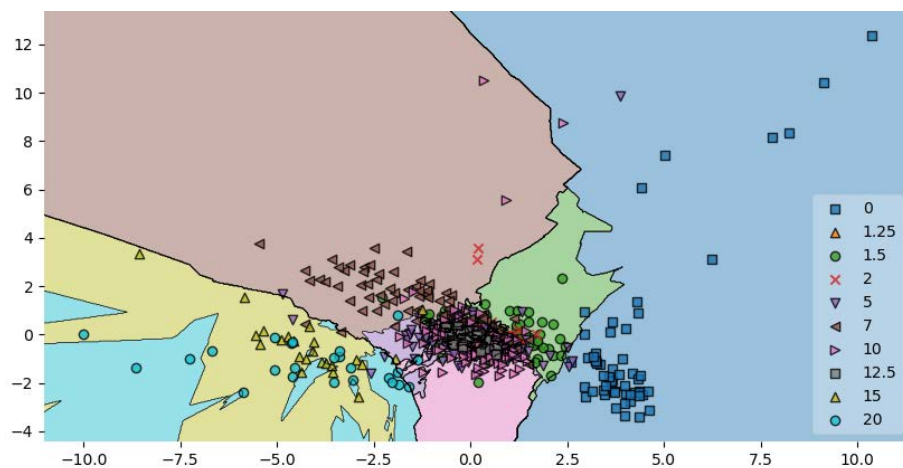
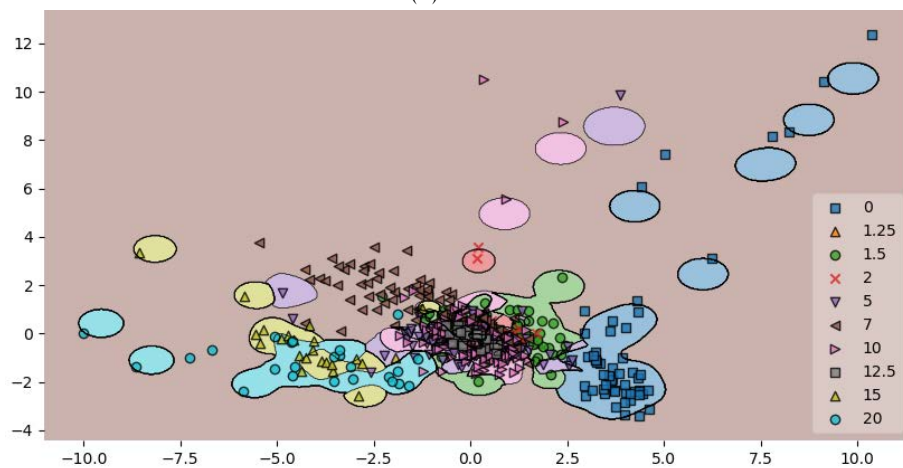
(c) *Random Forest*

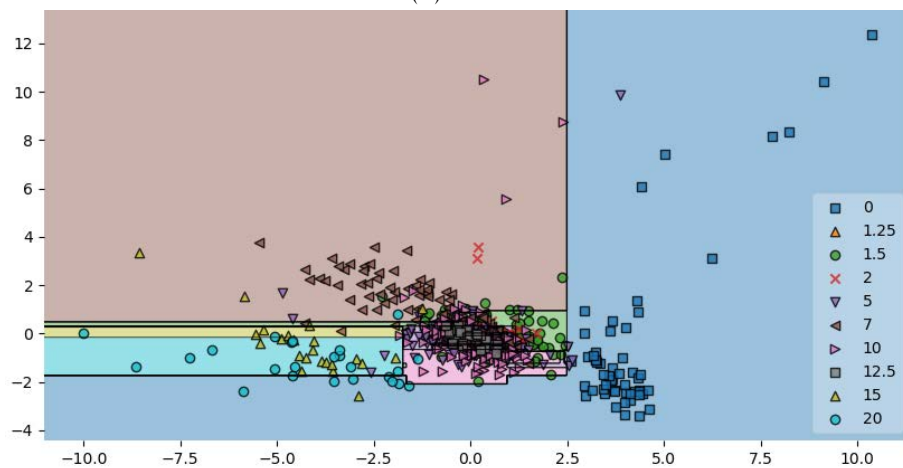
Figura F.6: Regiões de decisão para as concentrações de glucose (mM) dos três algoritmos estudados, baseados na média dos componentes YUV.



(a) KNN

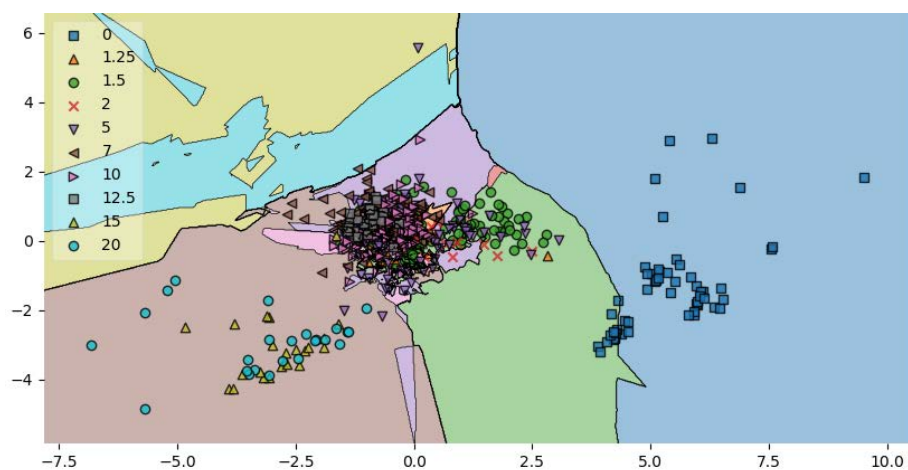


(b) SVM

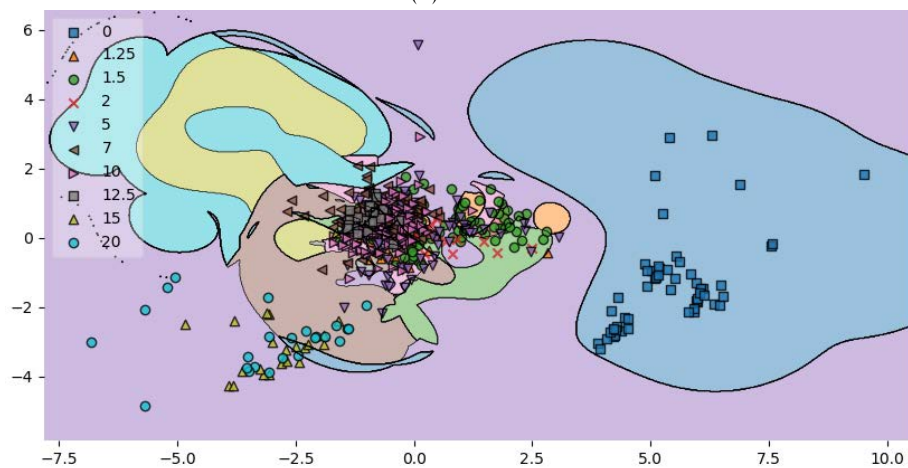


(c) *Random Forest*

Figura F.7: Regiões de decisão para as concentrações de glucose (mM) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes RGB.



(a) KNN



(b) SVM

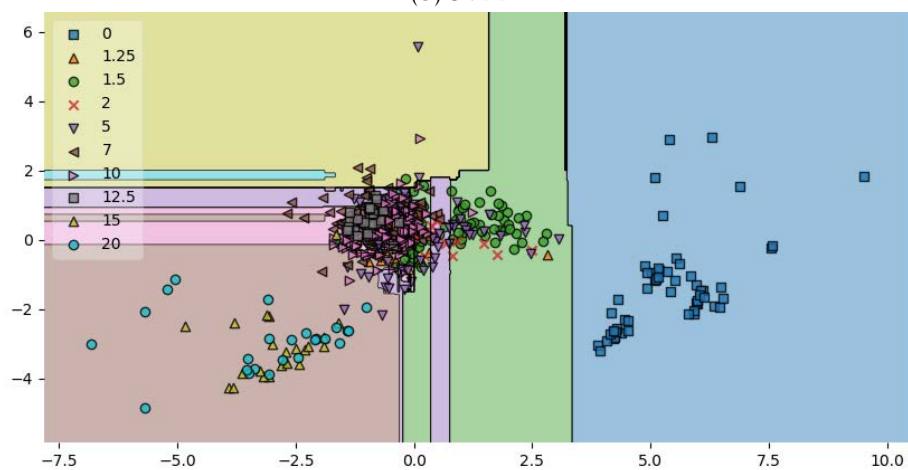
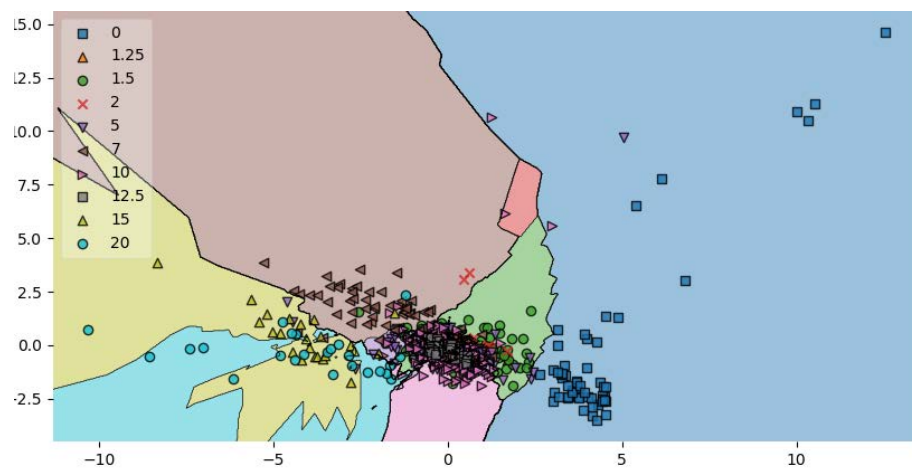
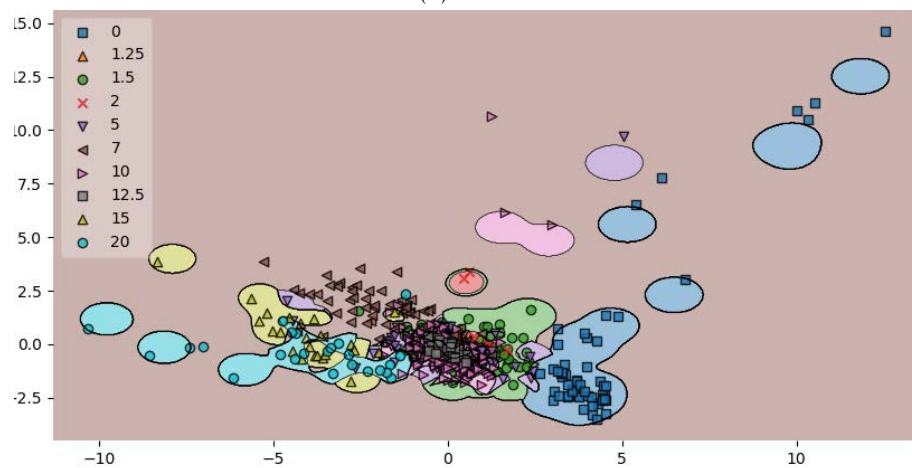
(c) *Random Forest*

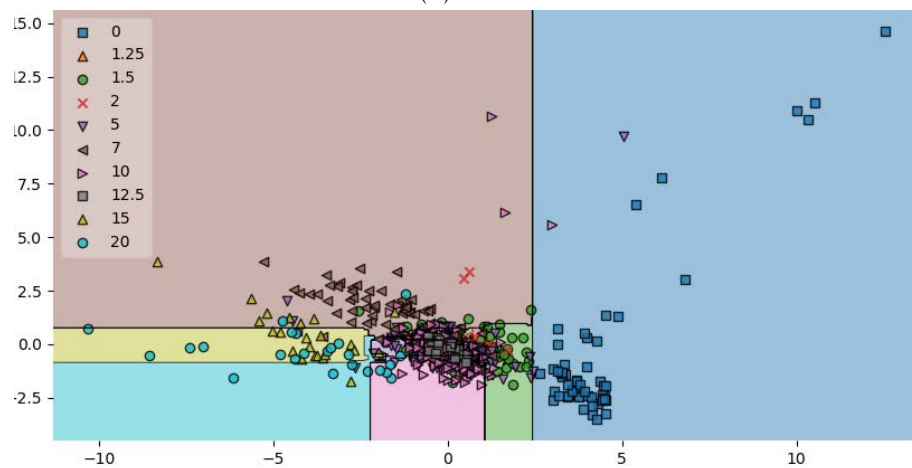
Figura F.8: Regiões de decisão para as concentrações de glucose (mM) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes HSV.



(a) KNN

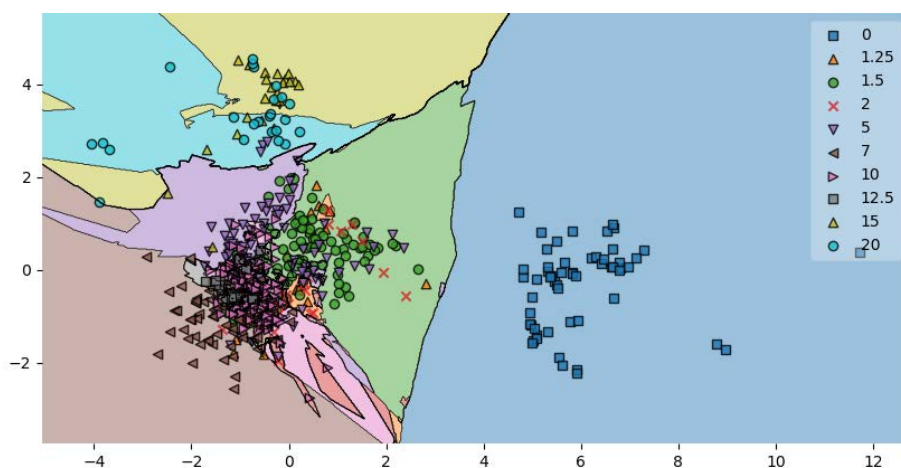


(b) SVM

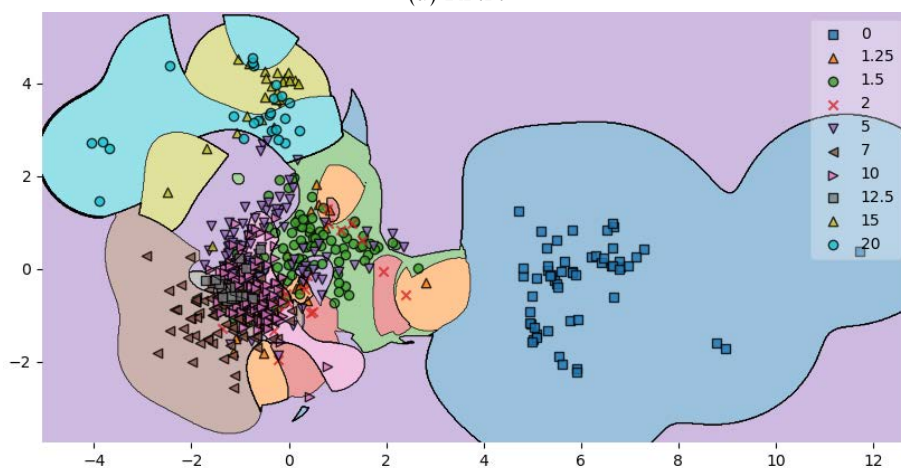


(c) *Random Forest*

Figura F.9: Regiões de decisão para as concentrações de glucose (mM) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes XYZ.



(a) KNN



(b) SVM

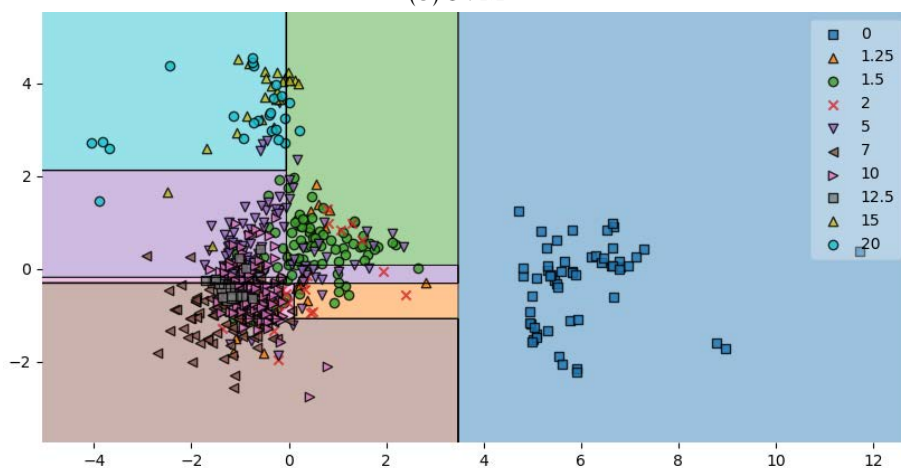
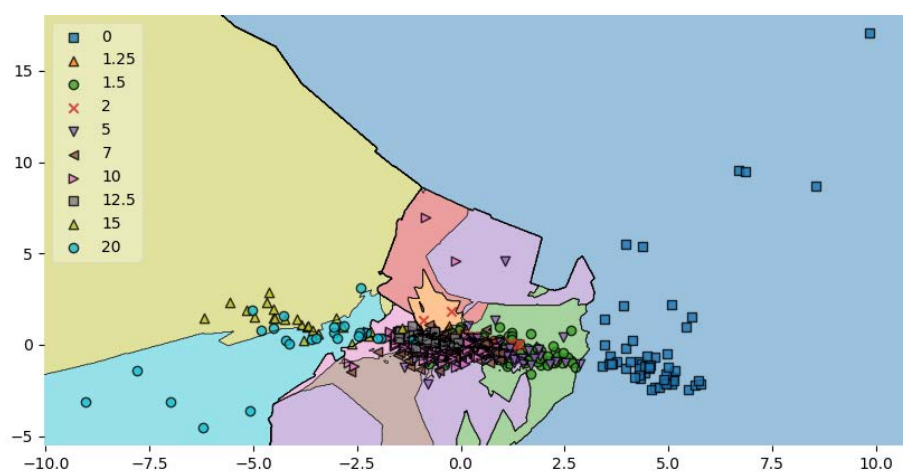
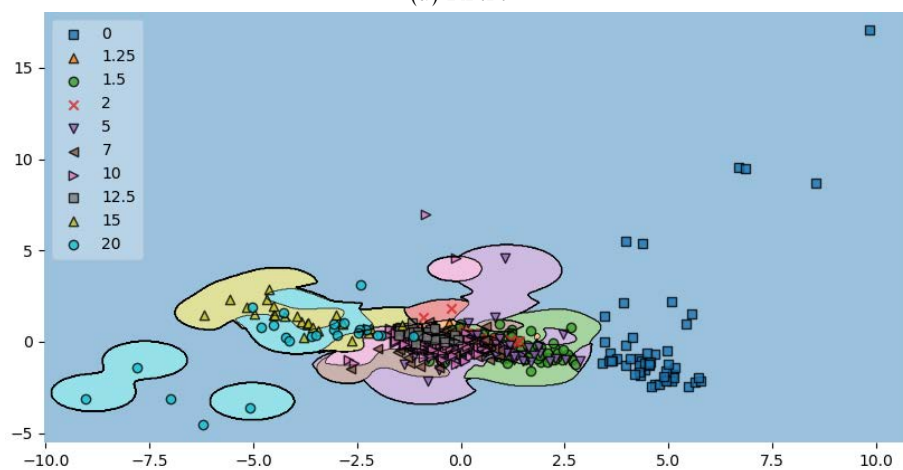
(c) *Random Forest*

Figura F.10: Regiões de decisão para as concentrações de glucose (mM) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes HLS.

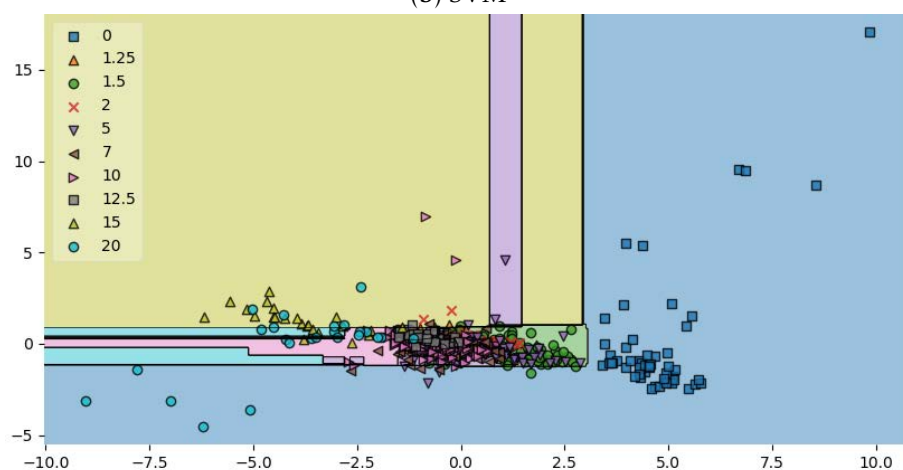




(a) KNN

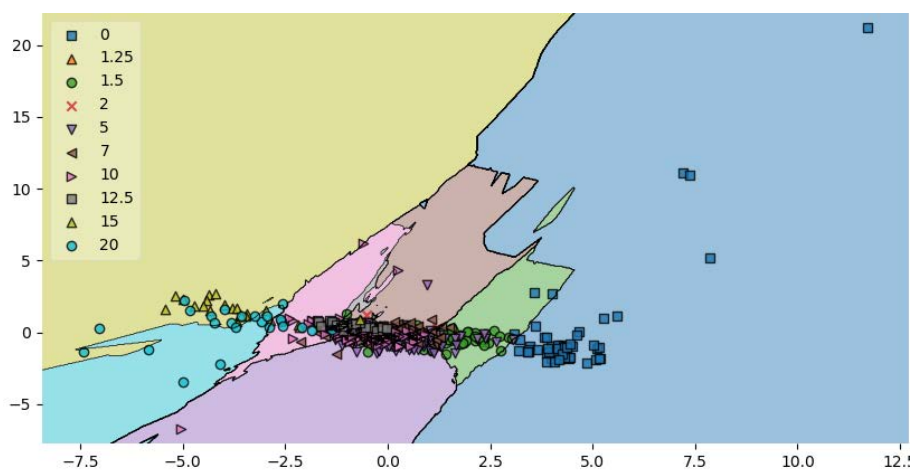


(b) SVM

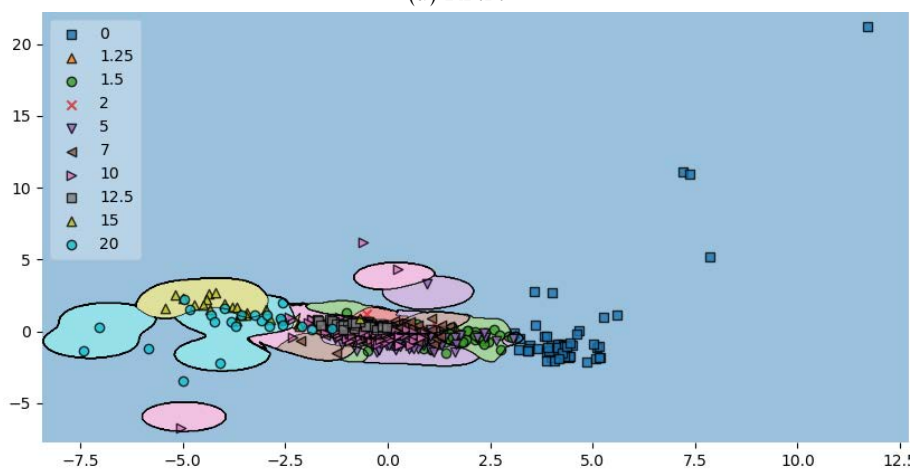


(c) *Random Forest*

Figura F.11: Regiões de decisão para as concentrações de glucose (mM) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes LAB.



(a) KNN



(b) SVM

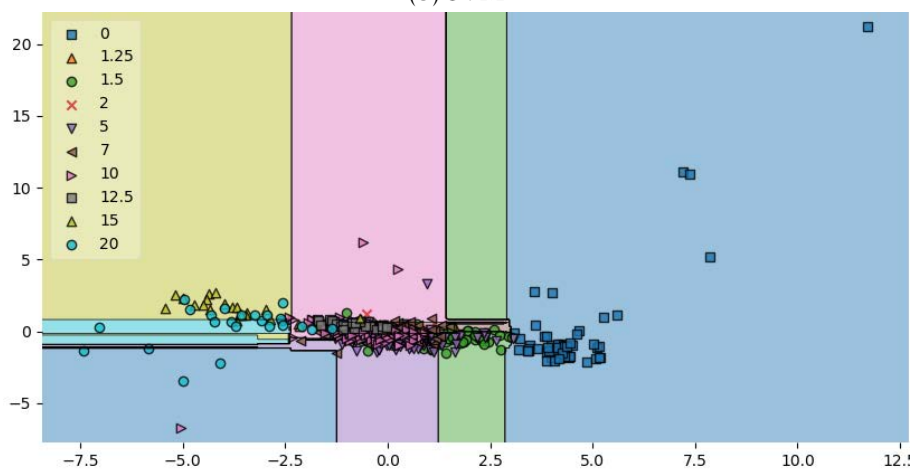
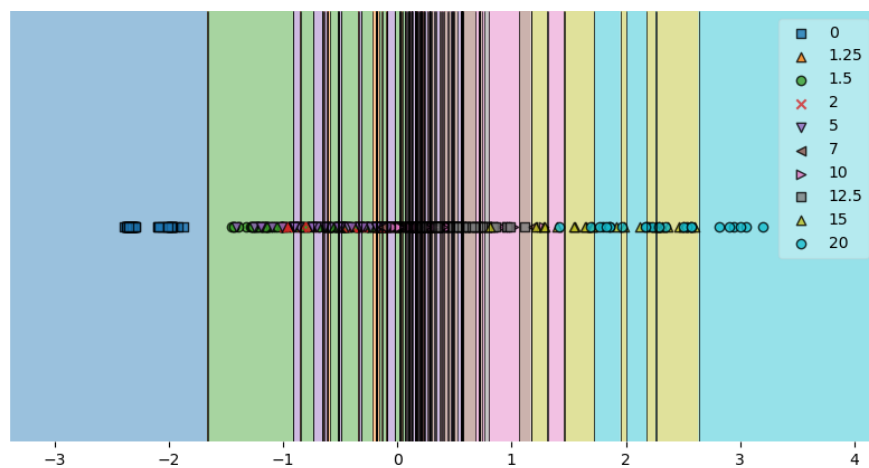
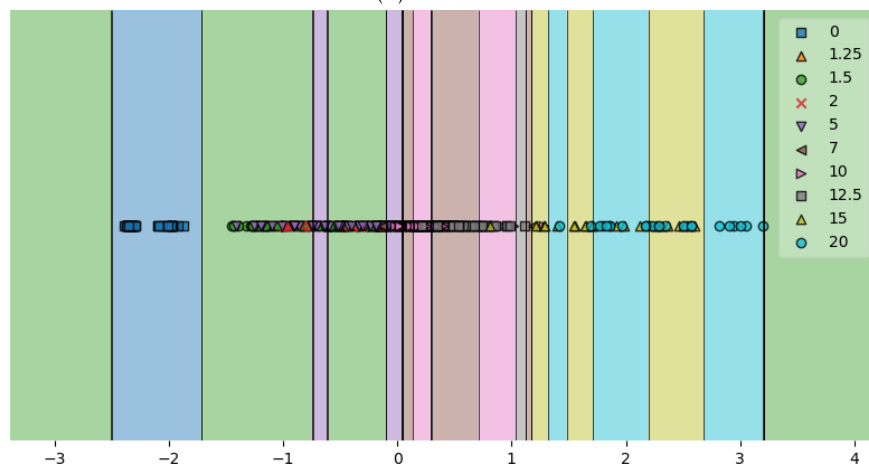
(c) *Random Forest*

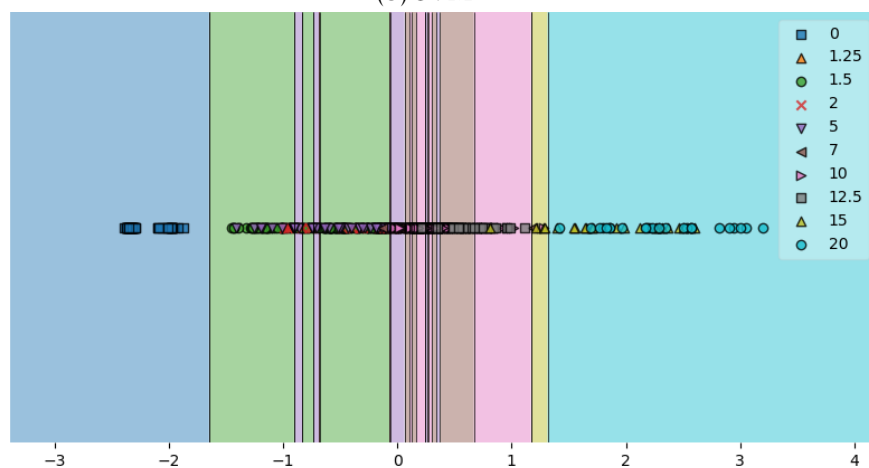
Figura F.12: Regiões de decisão para as concentrações de glucose (mM) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes YUV.



(a) KNN



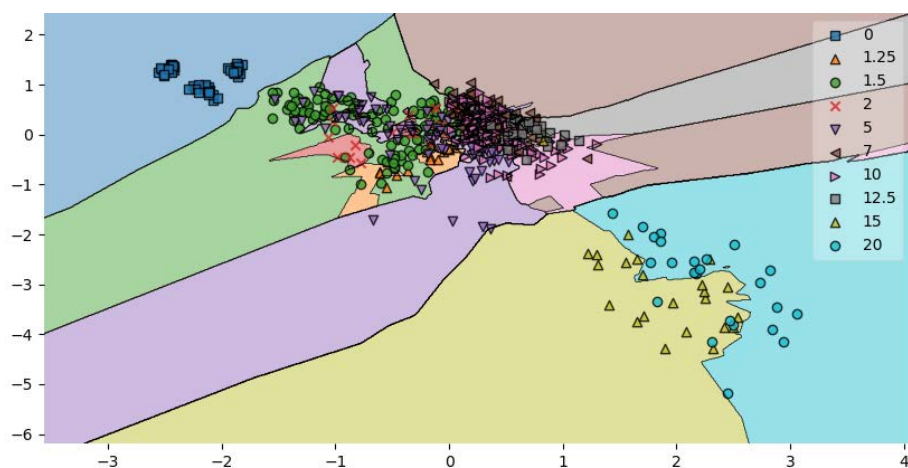
(b) SVM



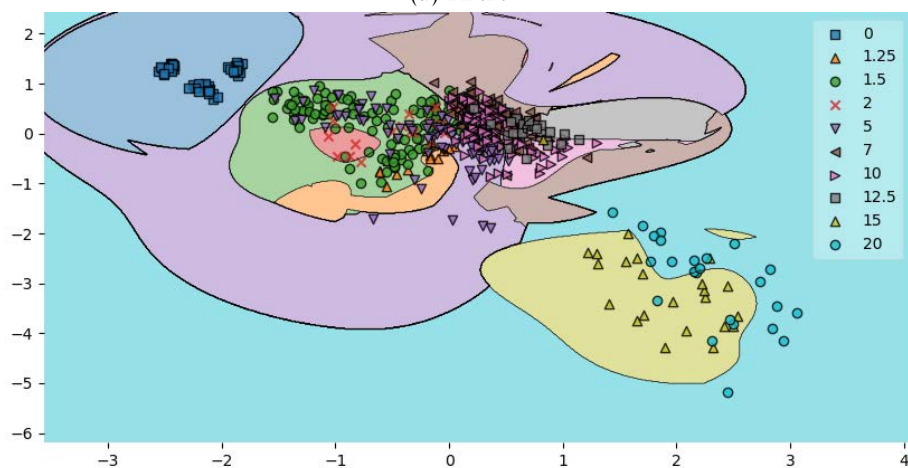
(c) *Random Forest*

Figura F.13: Regiões de decisão para as concentrações de glucose (mM) dos três algoritmos estudados, baseados no rácio entre os componentes vermelho e verde do espaço RGB.

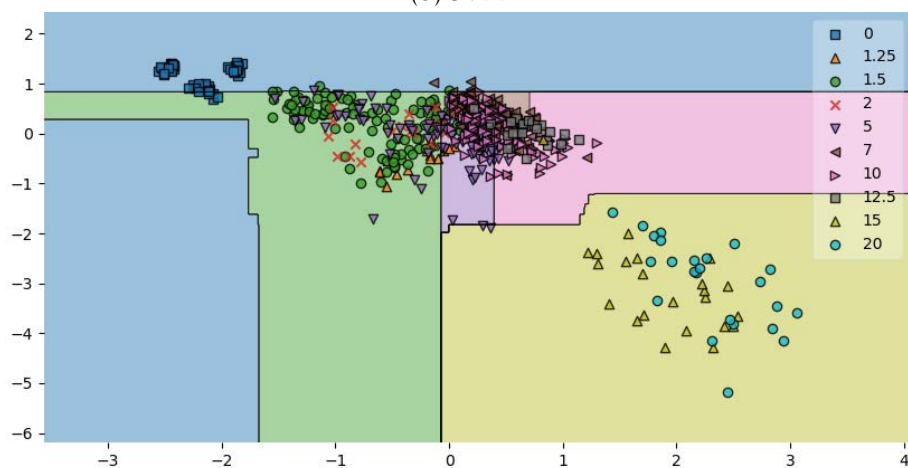




(a) KNN

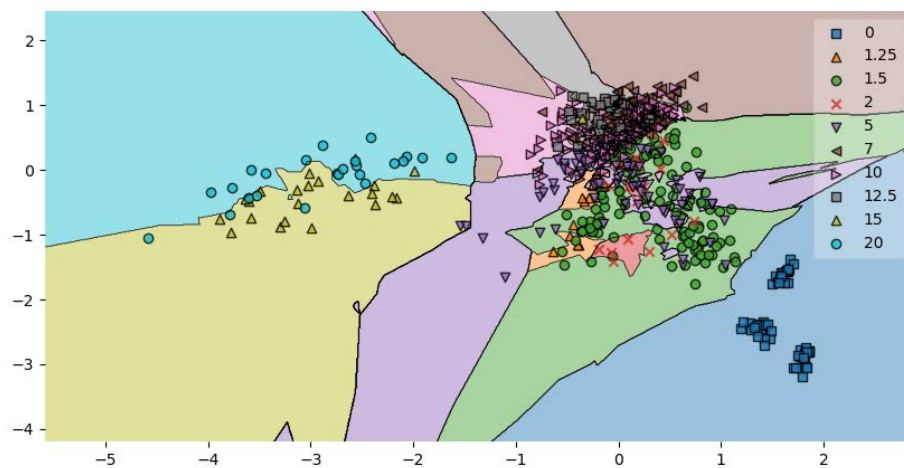


(b) SVM

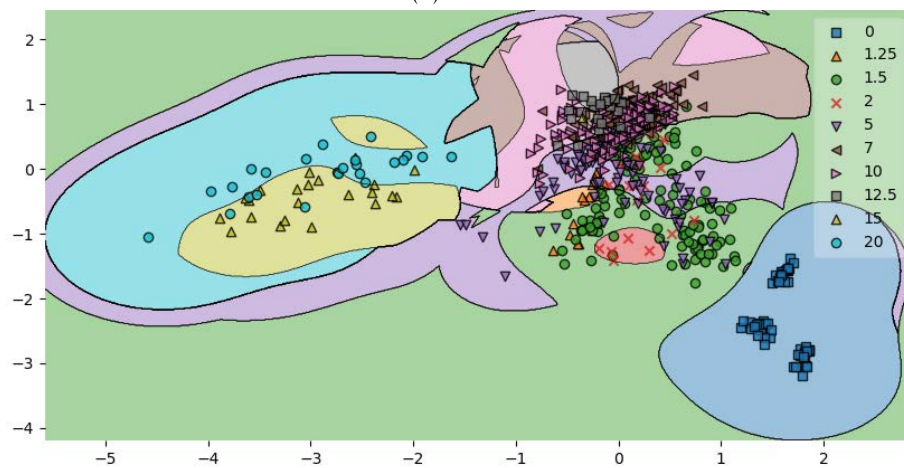


(c) Random Forest

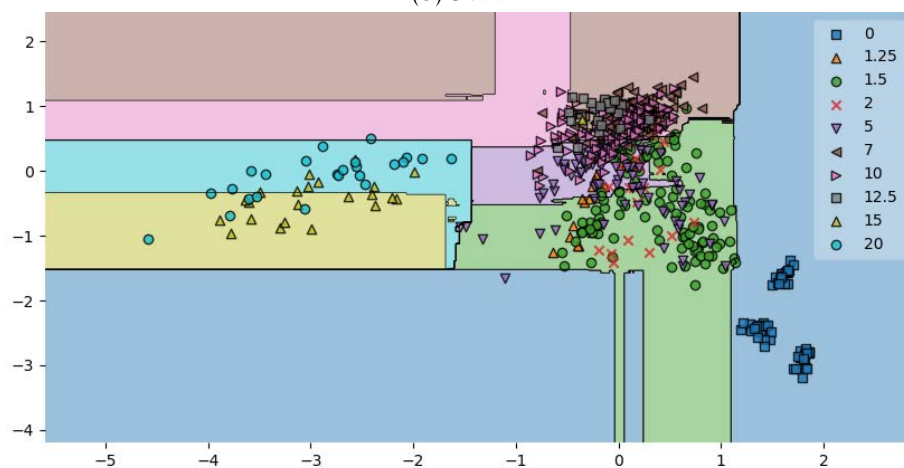
Figura F.14: Regiões de decisão para as concentrações de glucose (mM) dos três algoritmos estudados, baseados nos componentes S e V do espaço HSV.



(a) KNN



(b) SVM

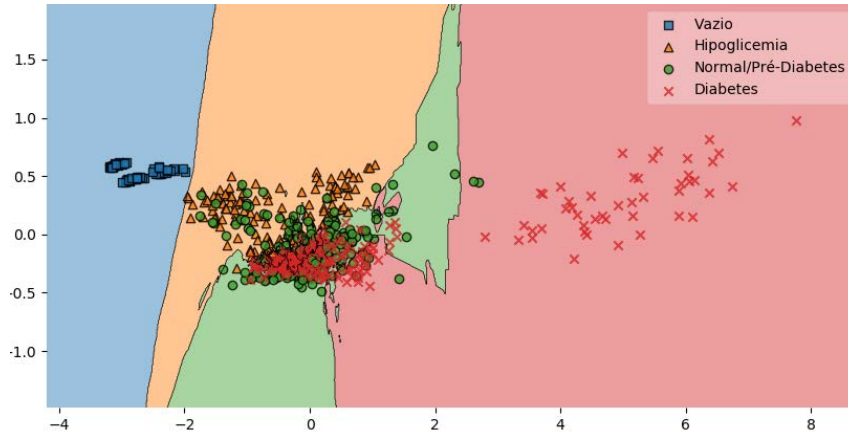


(c) *Random Forest*

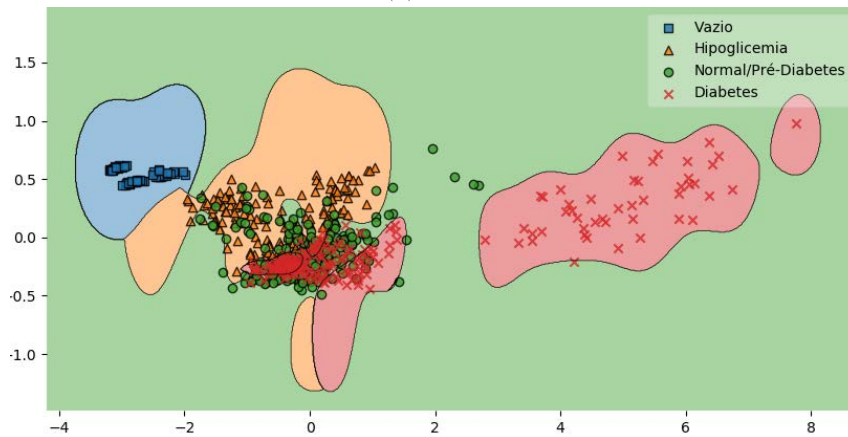
Figura F.15: Regiões de decisão para as concentrações de glucose (mM) dos três algoritmos estudados, baseados nos componentes L e S do espaço HLS.

## F.2 Quatro Níveis de Glucose

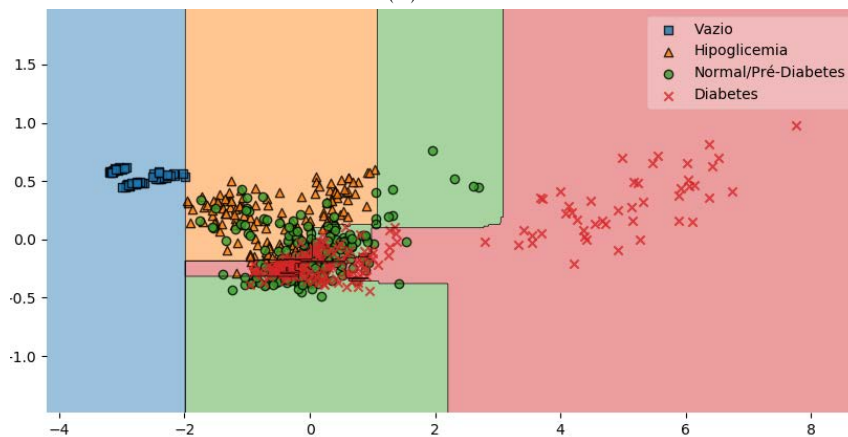
### F.2.1 Estado de jejum



(a) KNN



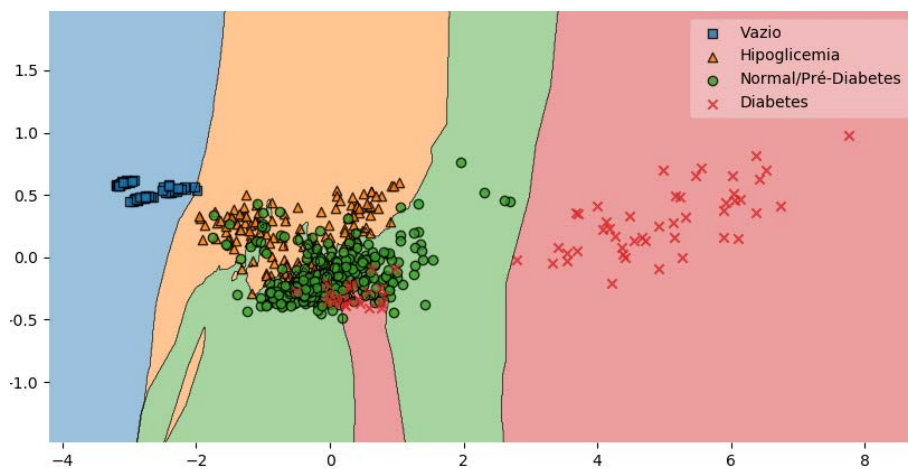
(b) SVM



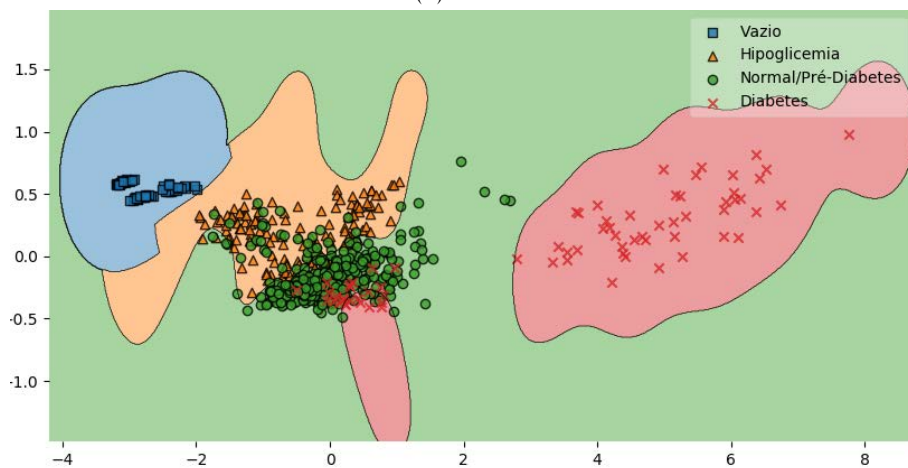
(c) Random Forest

Figura F.16: Regiões de decisão para os quatro níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados na média dos componentes RGB.

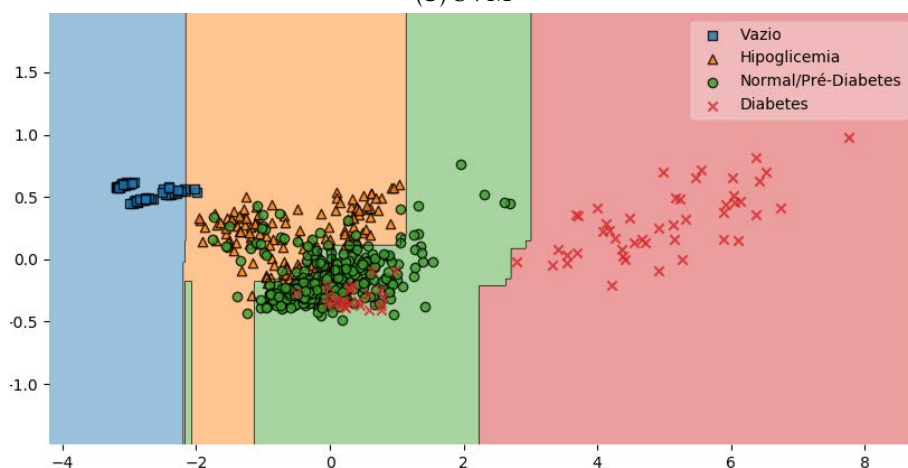
## F.2.2 Estado de duas horas após a refeição



(a) KNN

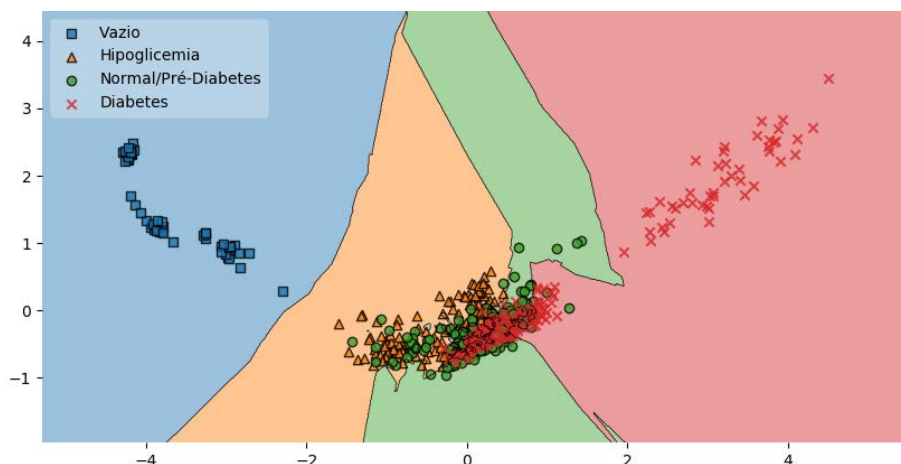


(b) SVM

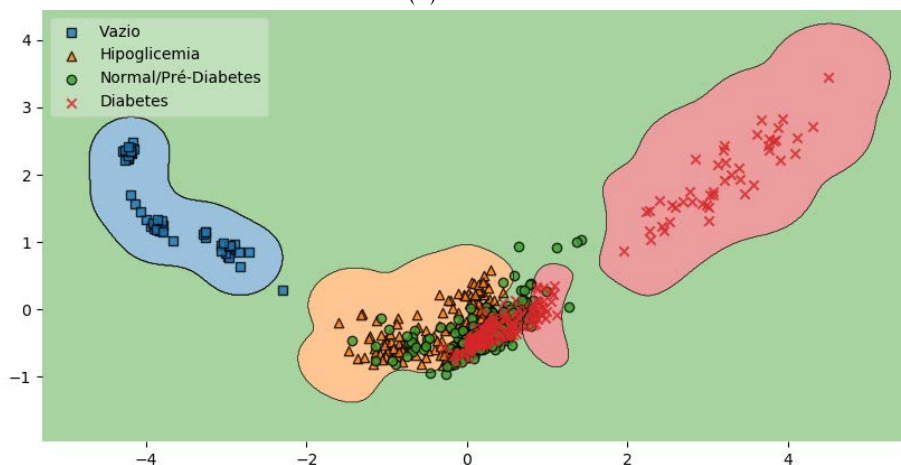


(c) Random Forest

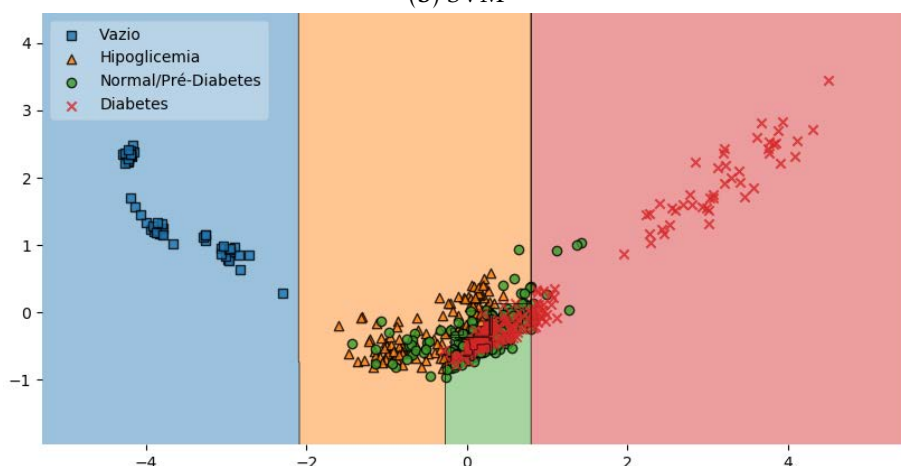
Figura F.31: Regiões de decisão para os quatro níveis de referência da glicose (duas horas após a refeição) dos três algoritmos estudados, baseados na média dos componentes RGB.



(a) KNN



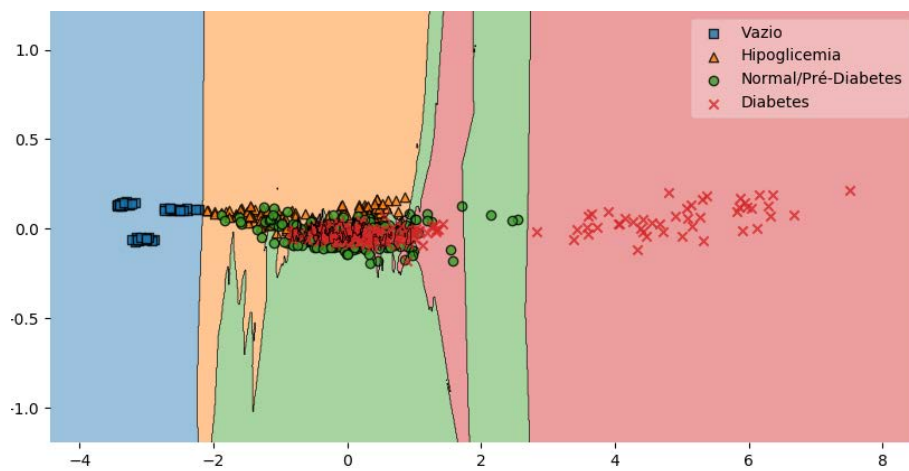
(b) SVM



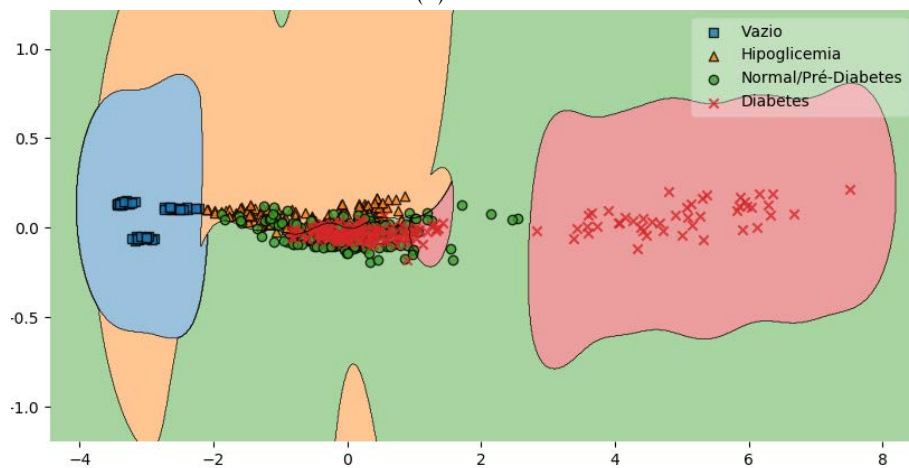
(c) Random Forest

Figura F.17: Regiões de decisão para os quatro níveis de referência da glicose (em jejum) dos três algoritmos estudados, baseados na média dos componentes HSV.

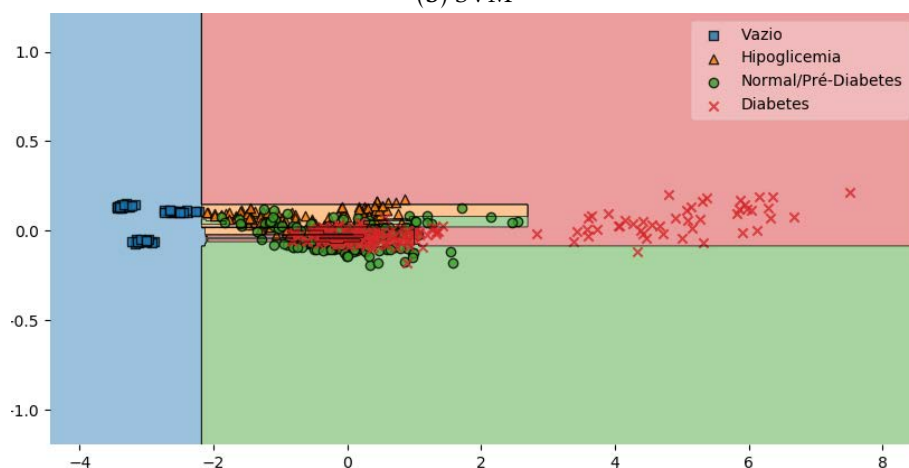




(a) KNN

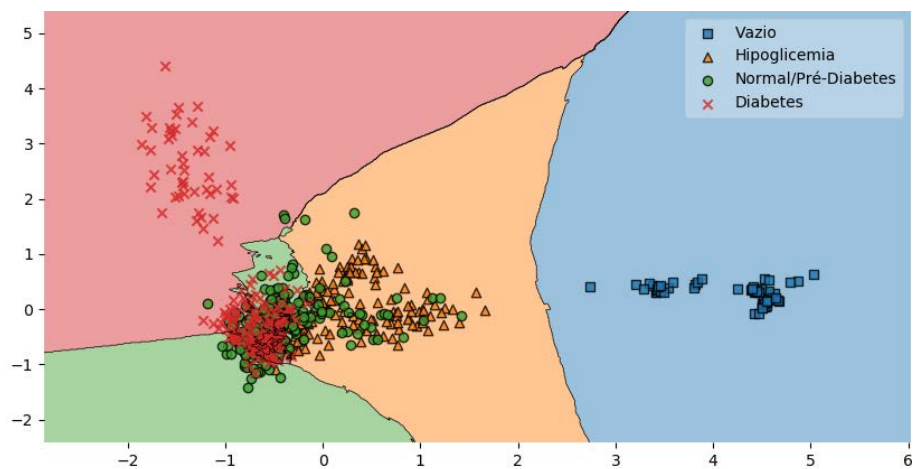


(b) SVM

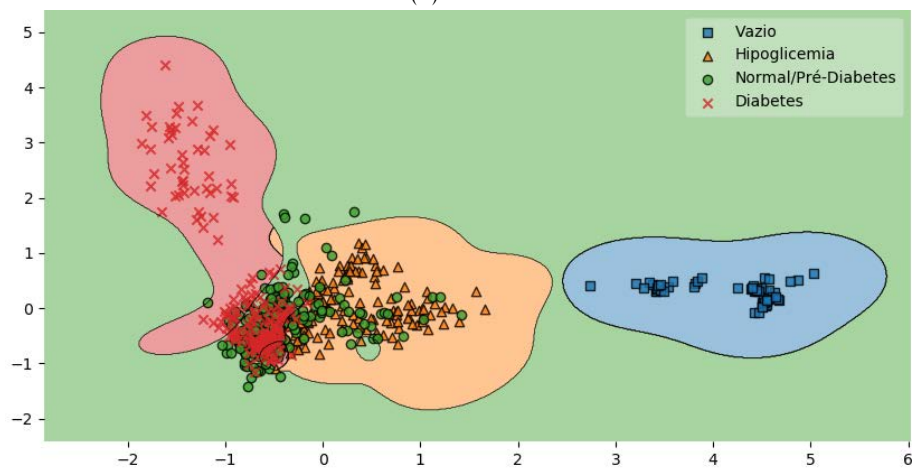


(c) Random Forest

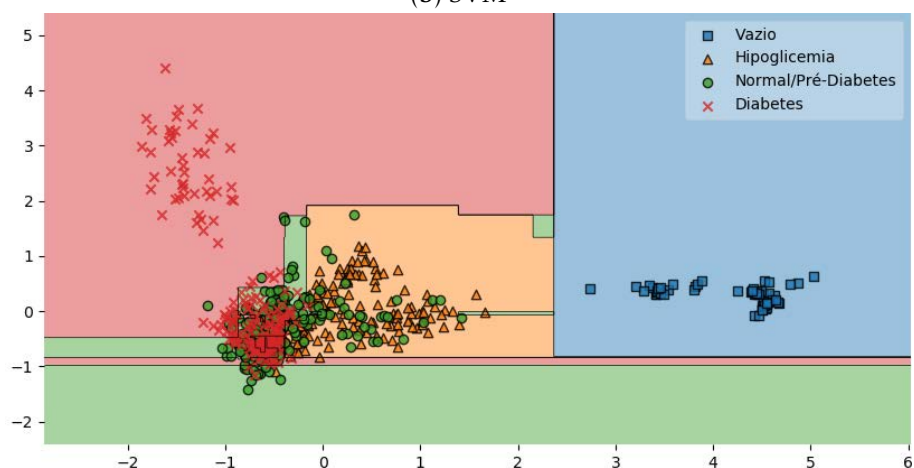
Figura F.18: Regiões de decisão para os quatro níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados na média dos componentes XYZ.



(a) KNN

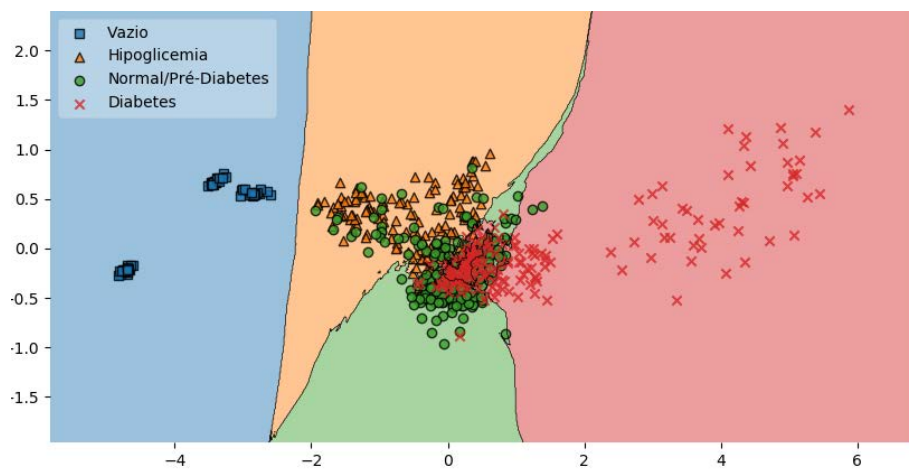


(b) SVM

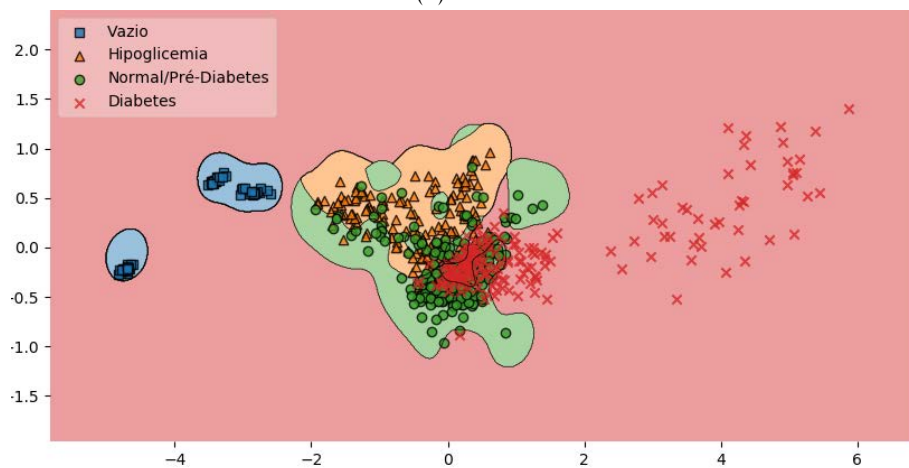


(c) *Random Forest*

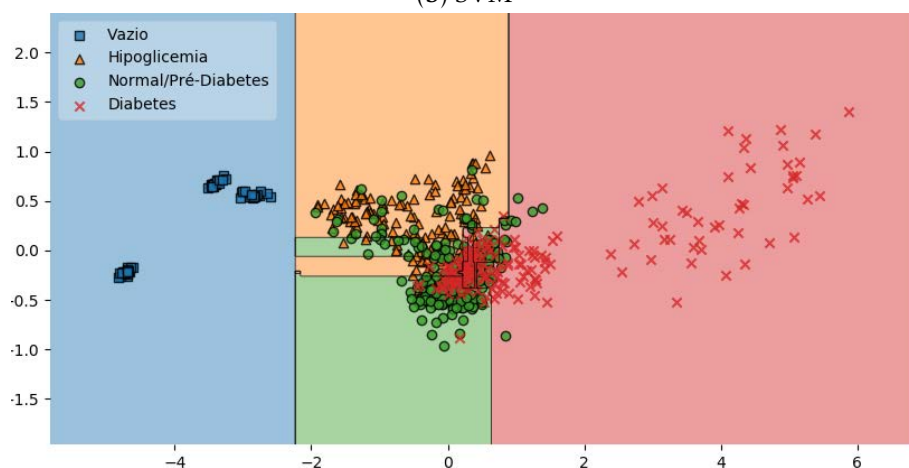
Figura F.19: Regiões de decisão para os quatro níveis de referência da glicose (em jejum) dos três algoritmos estudados, baseados na média dos componentes HLS.



(a) KNN



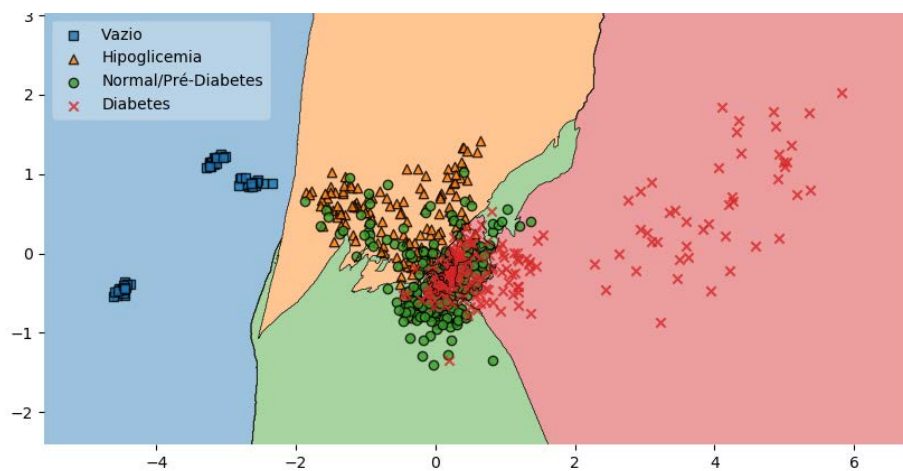
(b) SVM



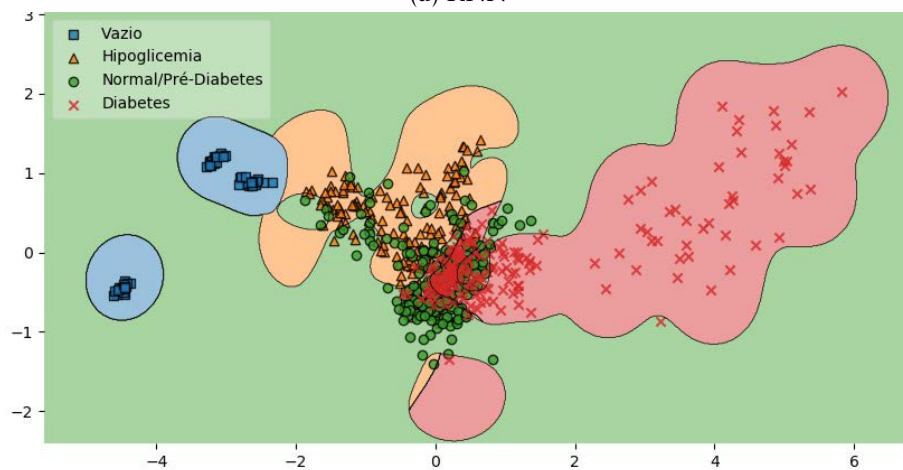
(c) Random Forest

Figura F.20: Regiões de decisão para os quatro níveis de referência da glicose (em jejum) dos três algoritmos estudados, baseados na média dos componentes LAB.

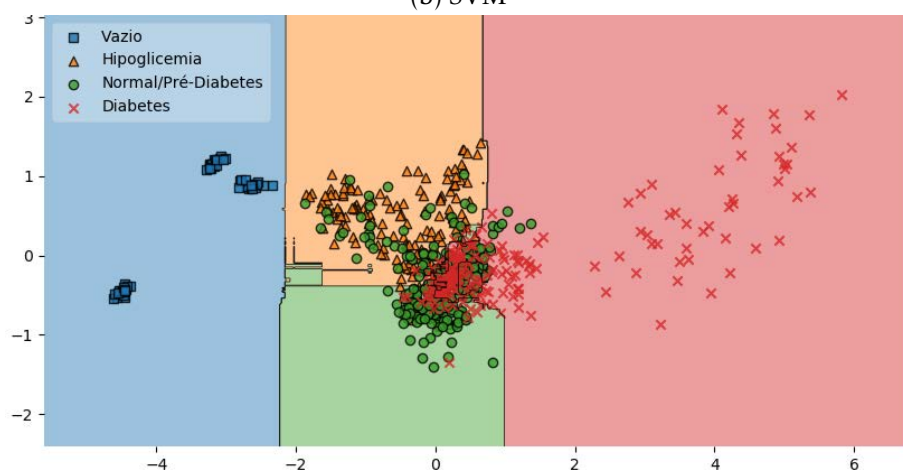




(a) KNN

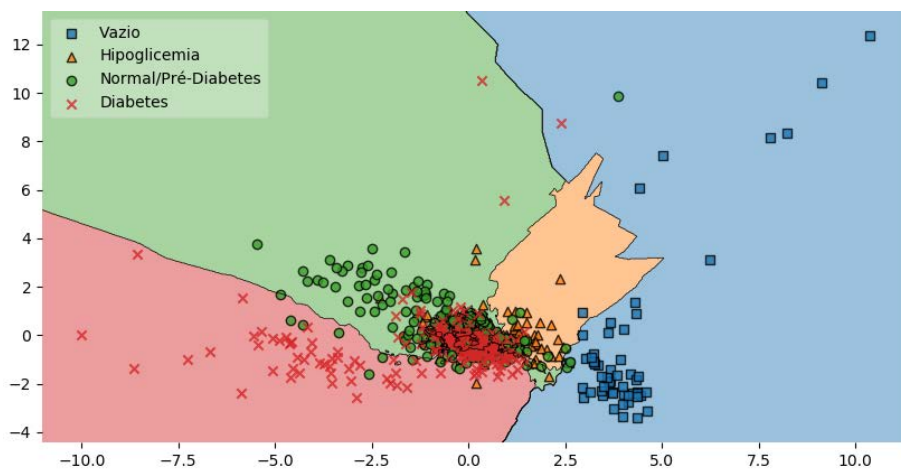


(b) SVM

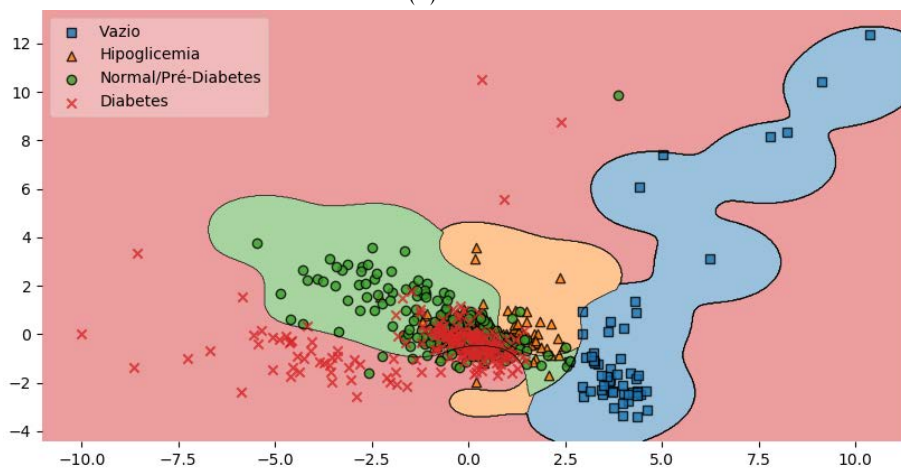


(c) Random Forest

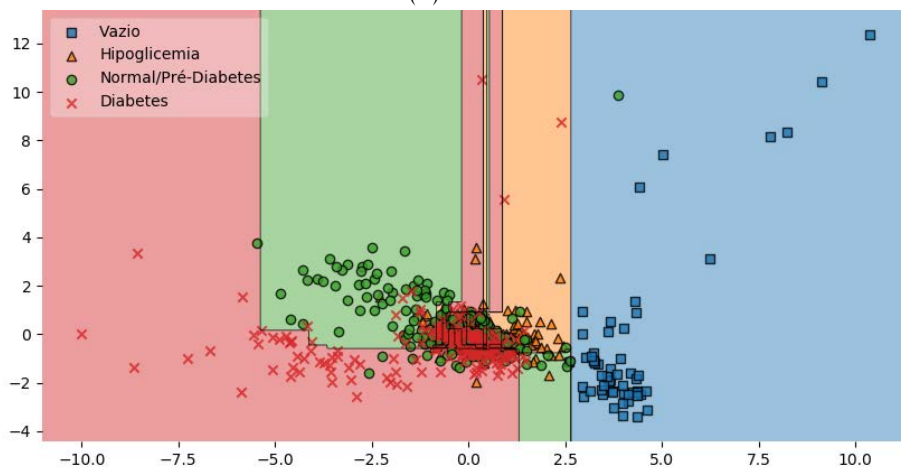
Figura F.21: Regiões de decisão para os quatro níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados na média dos componentes YUV.



(a) KNN

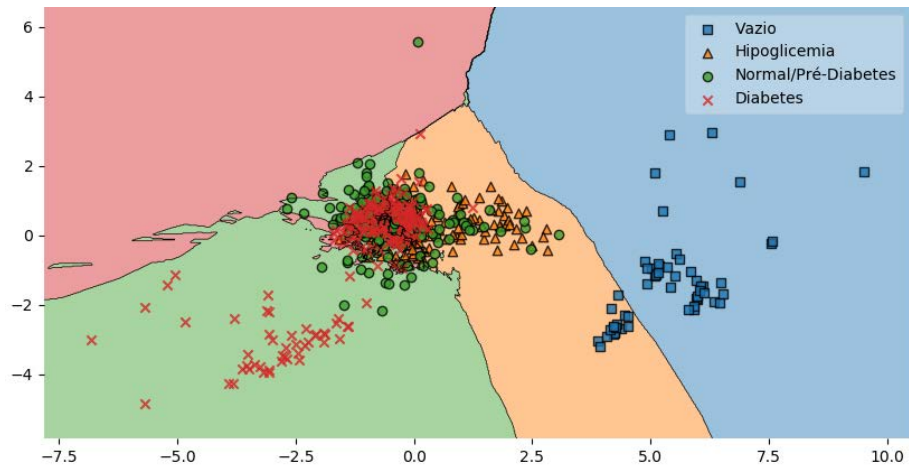


(b) SVM

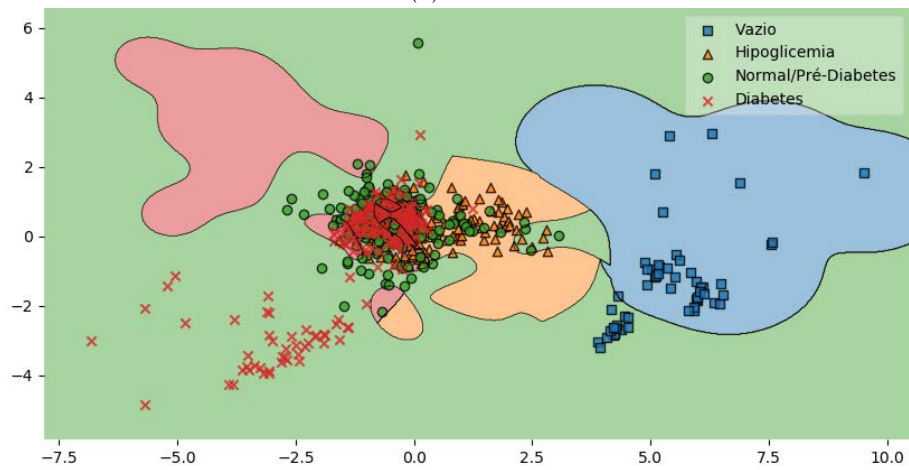


(c) Random Forest

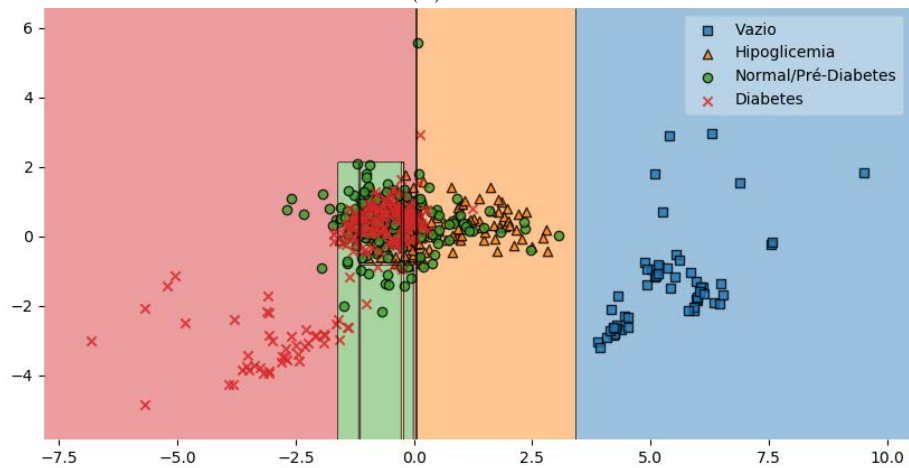
Figura F.22: Regiões de decisão para os quatro níveis de referência da glicose (em jejum) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes RGB.



(a) KNN

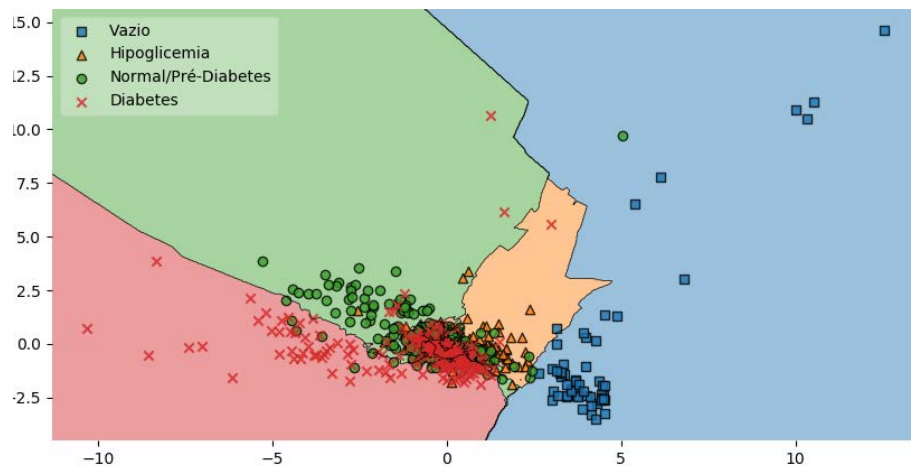


(b) SVM

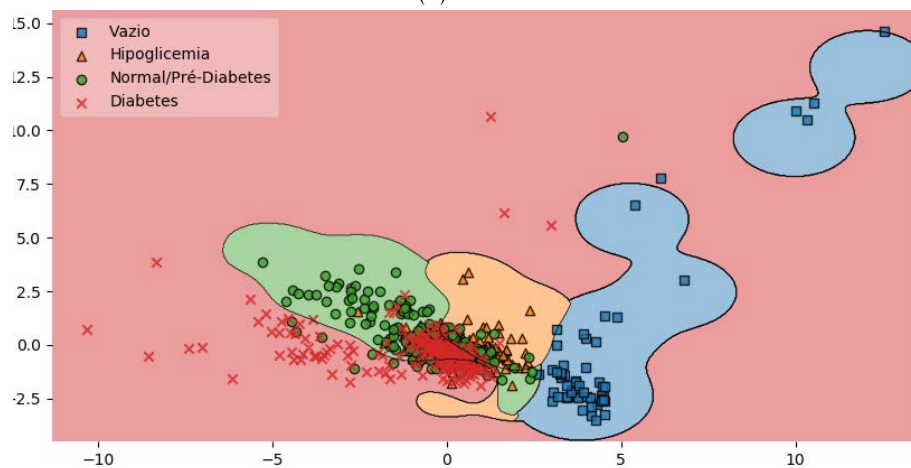


(c) Random Forest

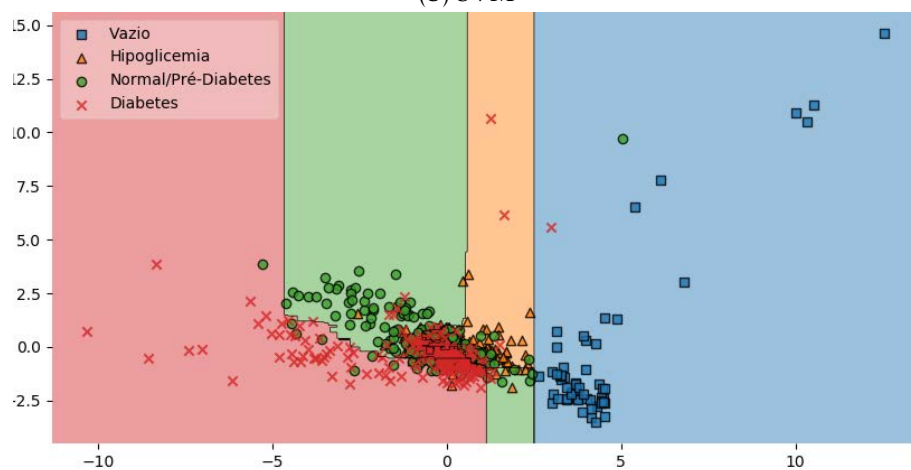
Figura F.23: Regiões de decisão para os quatro níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes HSV.



(a) KNN

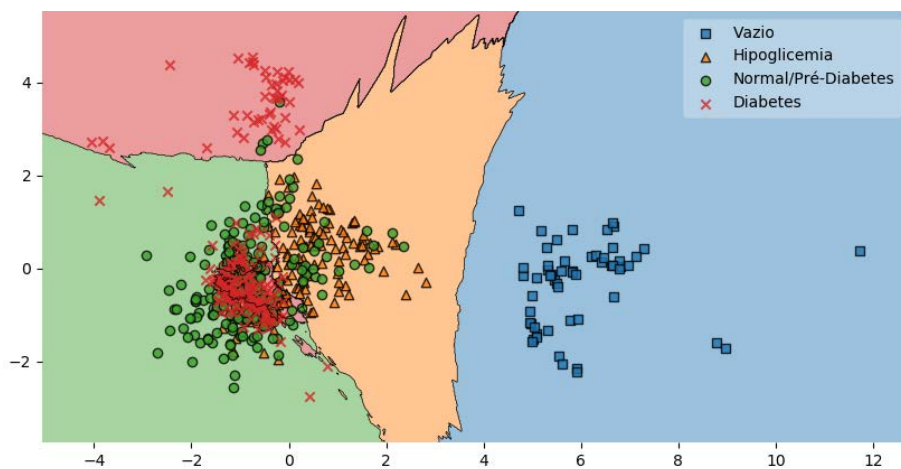


(b) SVM

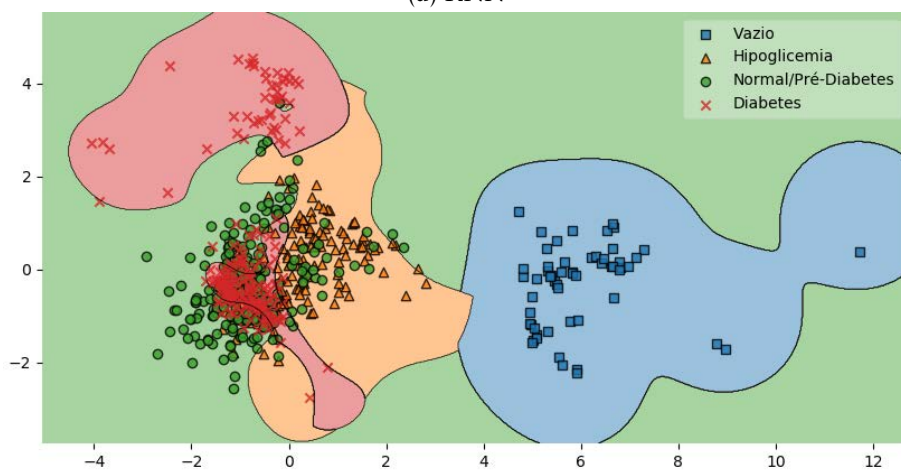


(c) Random Forest

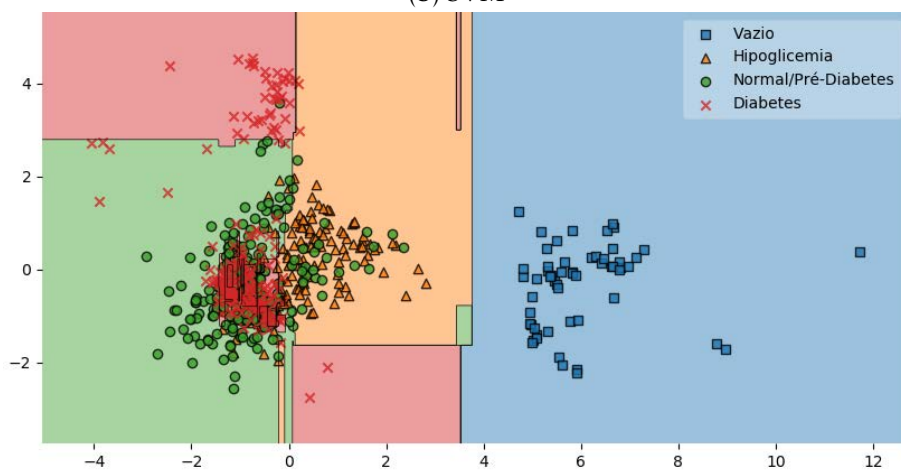
Figura F.24: Regiões de decisão para os quatro níveis de referência da glicose (em jejum) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes XYZ.



(a) KNN



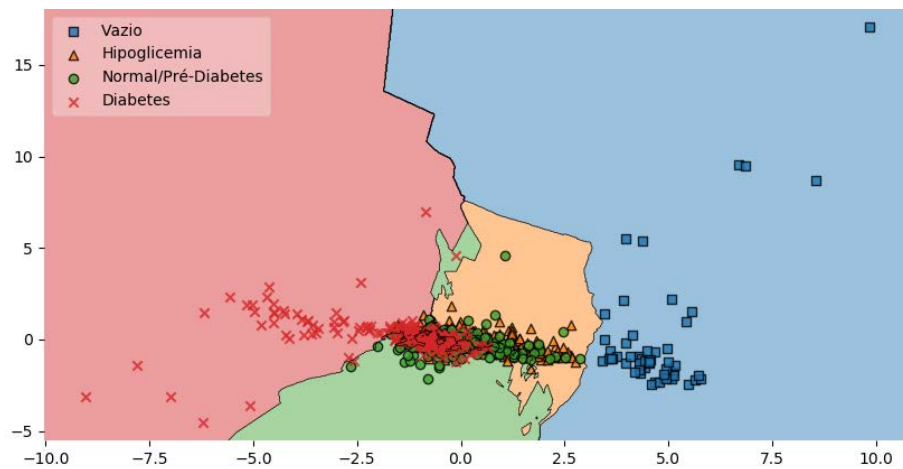
(b) SVM



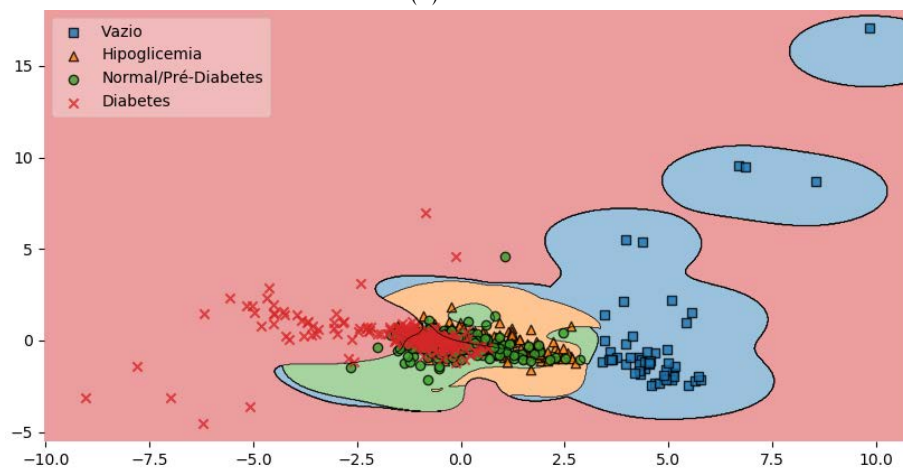
(c) Random Forest

Figura F.25: Regiões de decisão para os quatro níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes HLS.

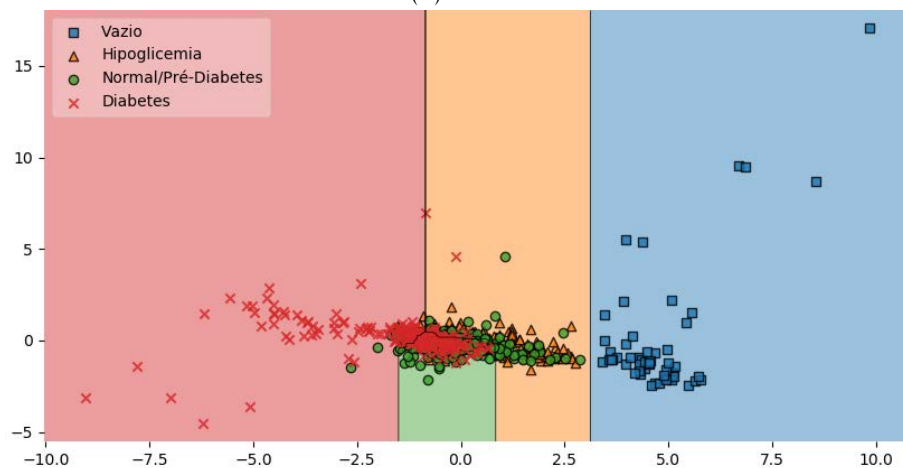




(a) KNN

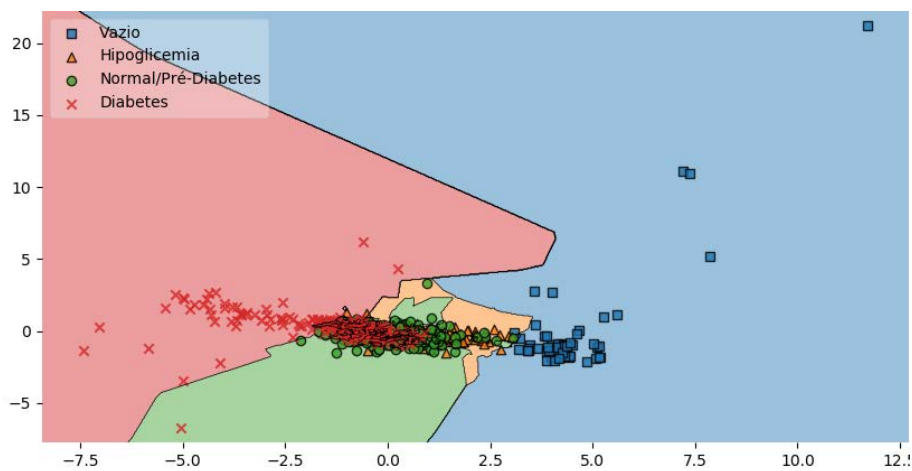


(b) SVM

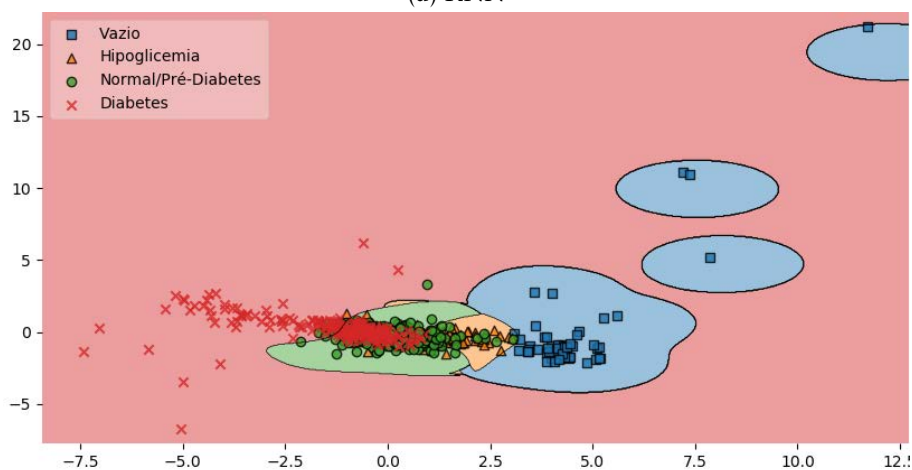


(c) *Random Forest*

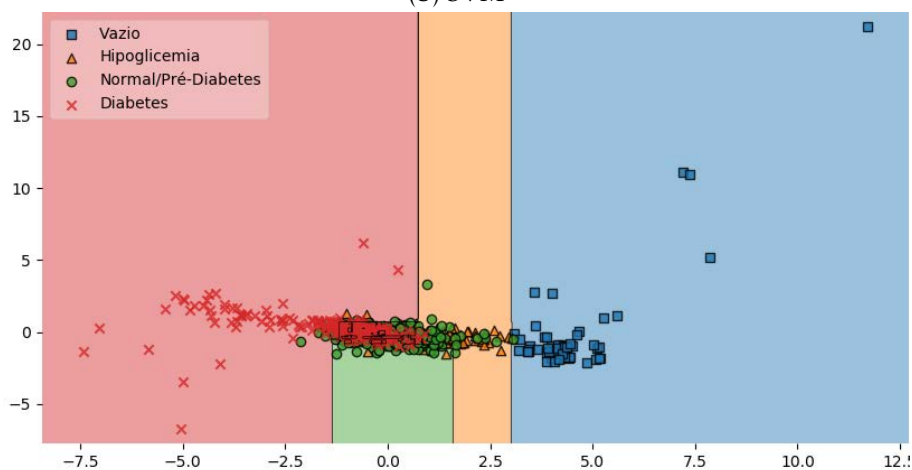
Figura F.26: Regiões de decisão para os quatro níveis de referência da glicose (em jejum) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes LAB.



(a) KNN

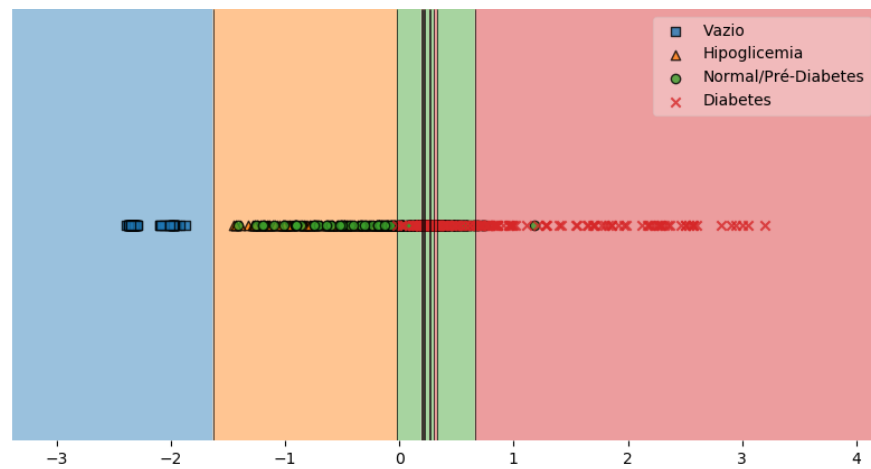


(b) SVM

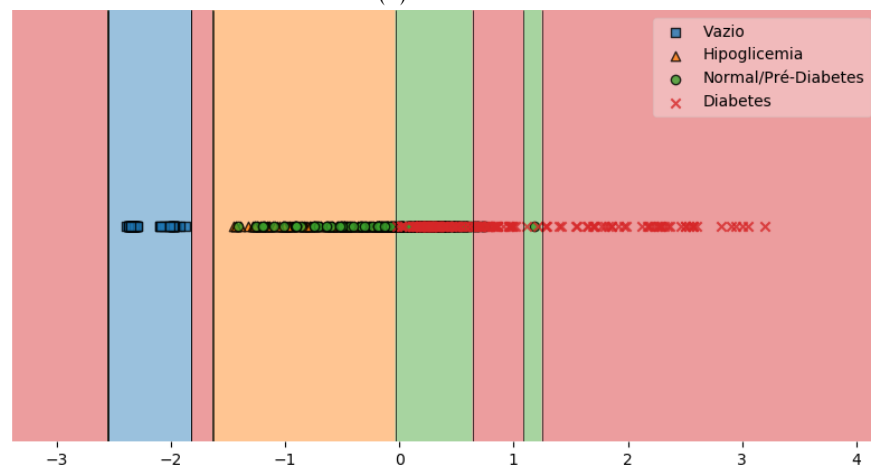


(c) Random Forest

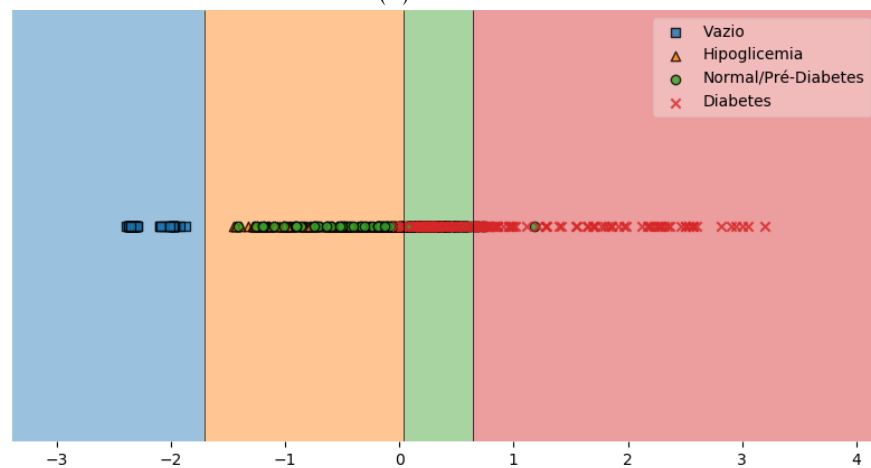
Figura F.27: Regiões de decisão para os quatro níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes YUV.



(a) KNN



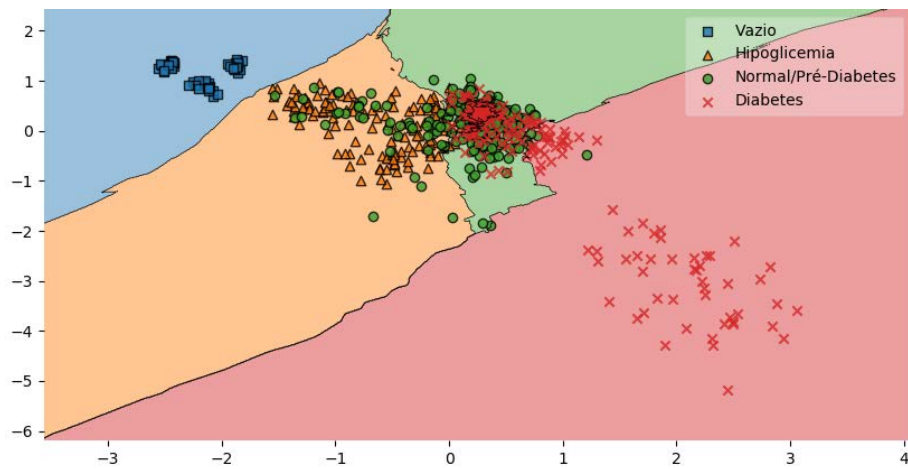
(b) SVM



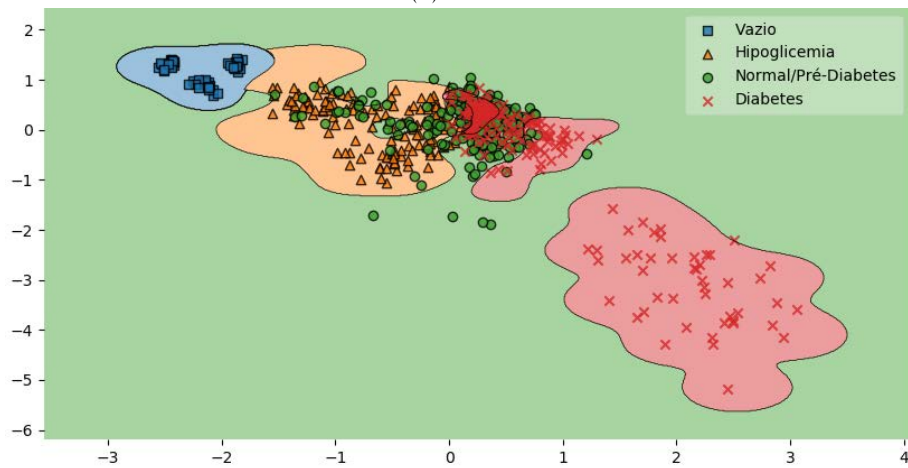
(c) Random Forest

Figura F.28: Regiões de decisão para os quatro níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados no rácio entre os componentes vermelho e verde do espaço RGB.

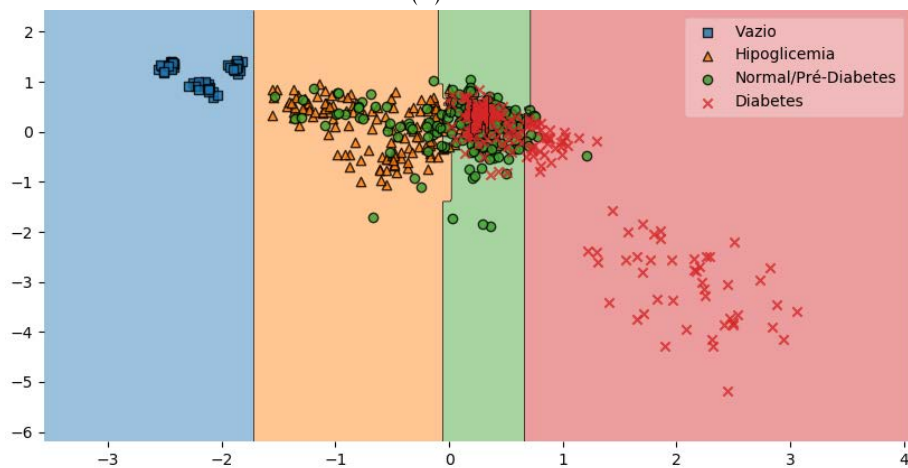




(a) KNN

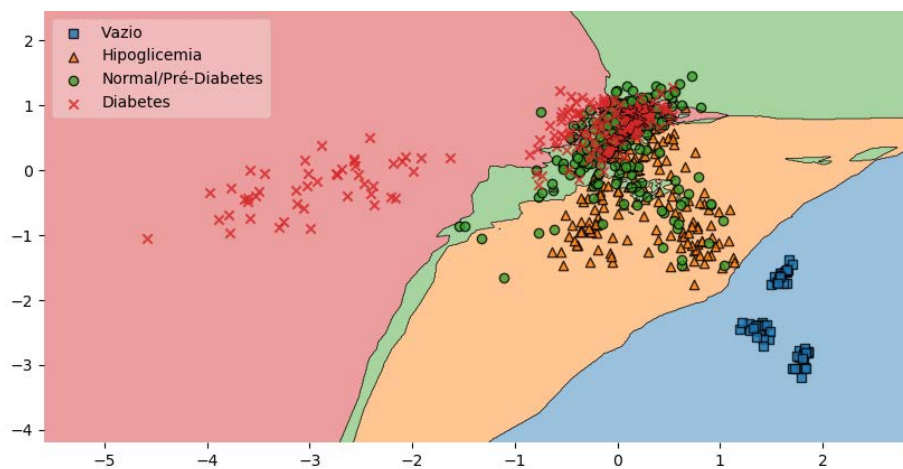


(b) SVM

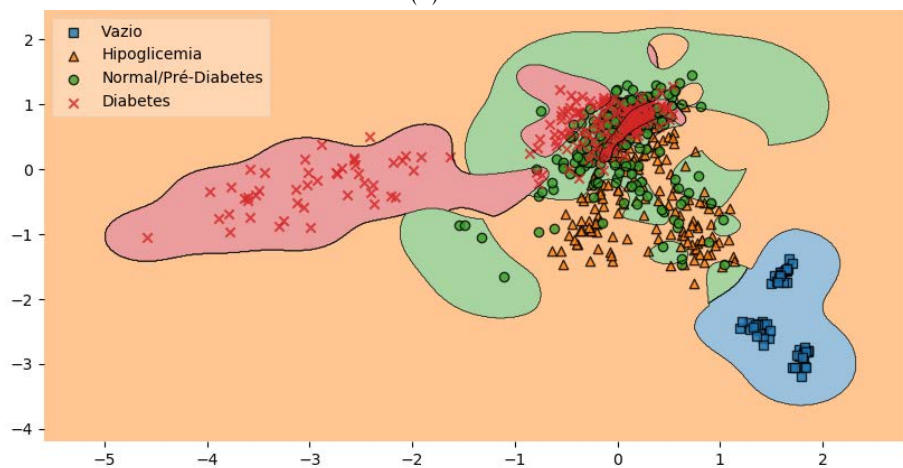


(c) *Random Forest*

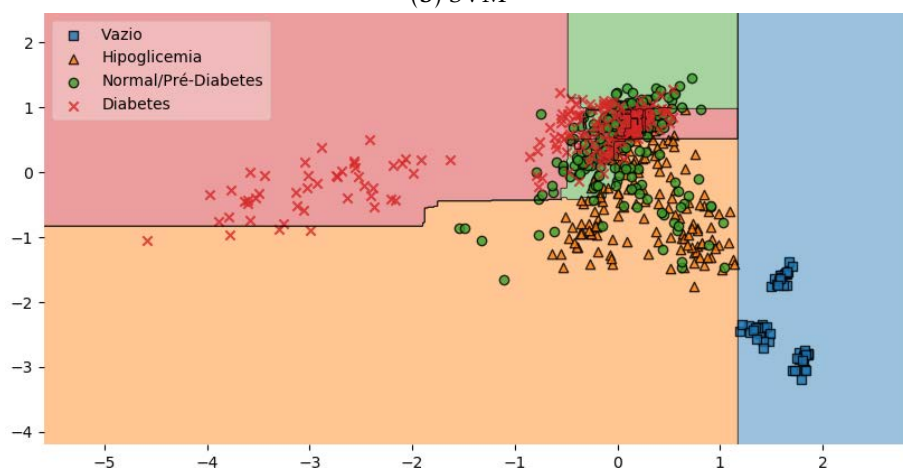
Figura F.29: Regiões de decisão para os quatro níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados nos componentes S e V do espaço HSV.



(a) KNN

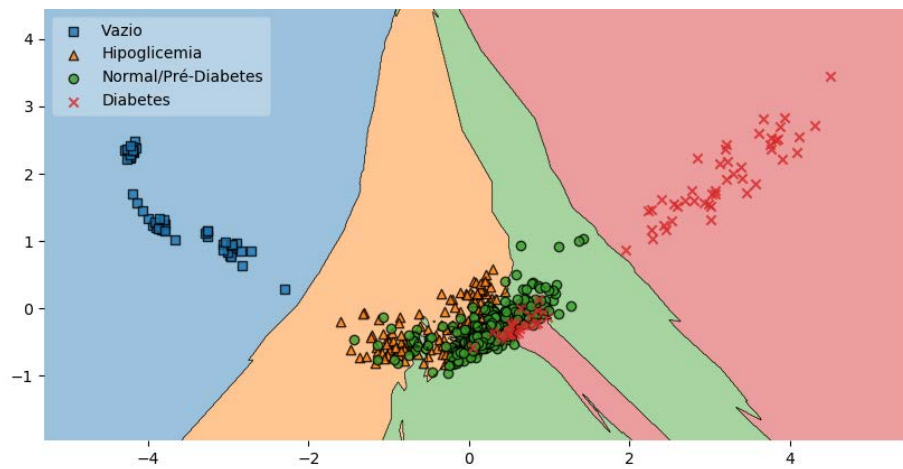


(b) SVM

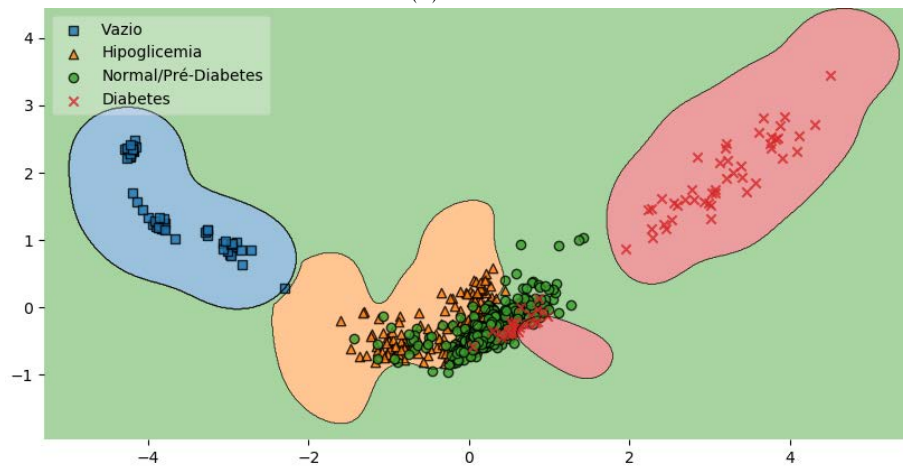


(c) Random Forest

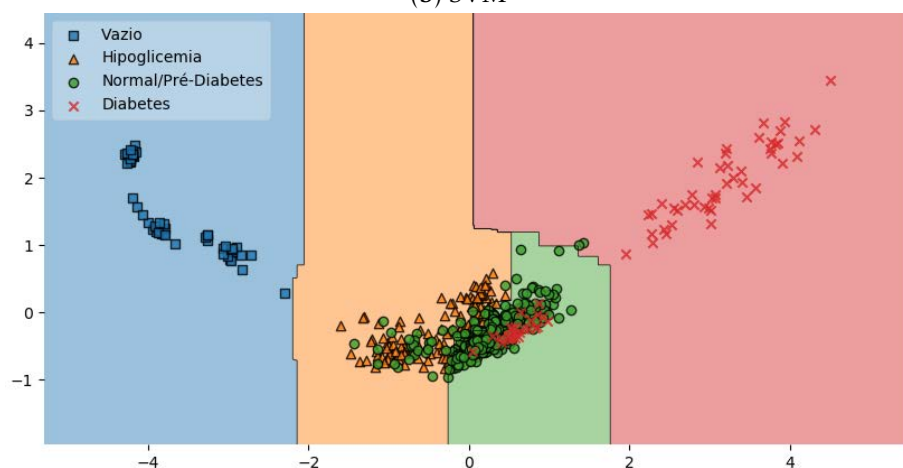
Figura F.30: Regiões de decisão para os quatro níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados nos componentes L e S do espaço HLS.



(a) KNN

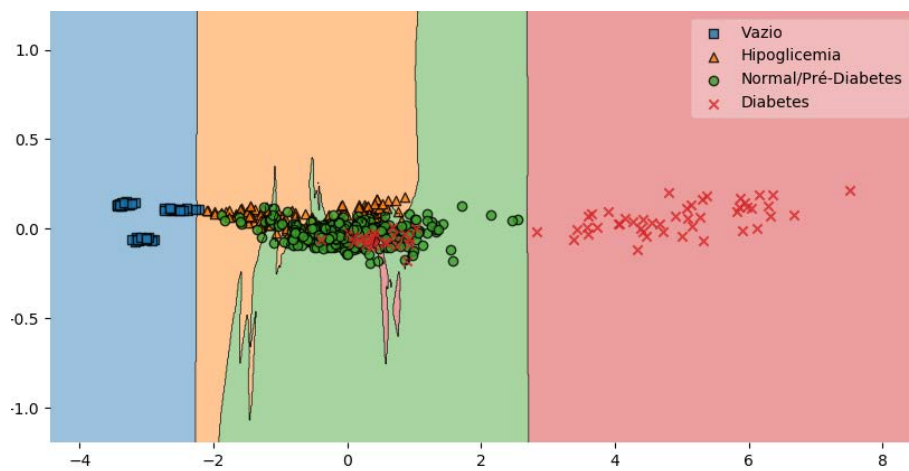


(b) SVM

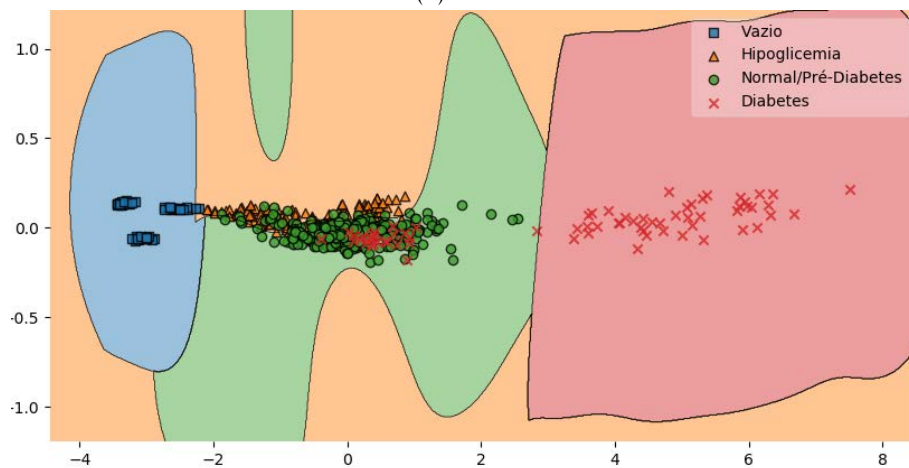


(c) Random Forest

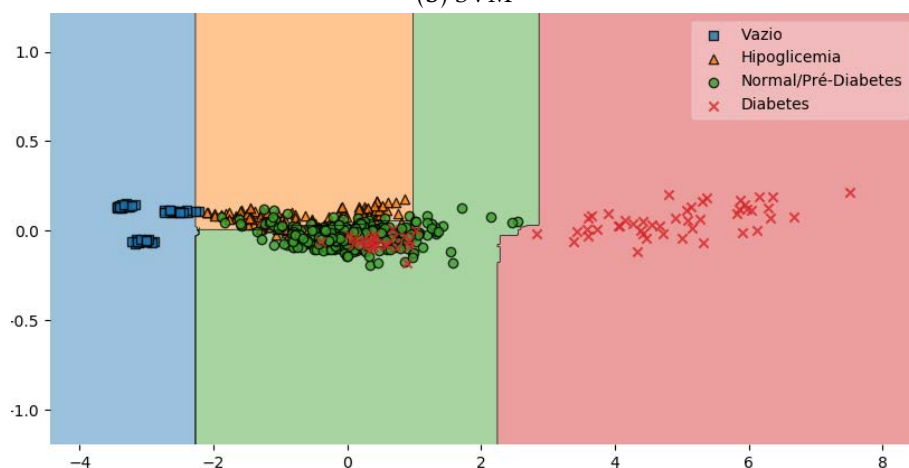
Figura F.32: Regiões de decisão para os quatro níveis de referência da glucose (duas horas após a refeição) dos três algoritmos estudados, baseados na média dos componentes HSV.



(a) KNN

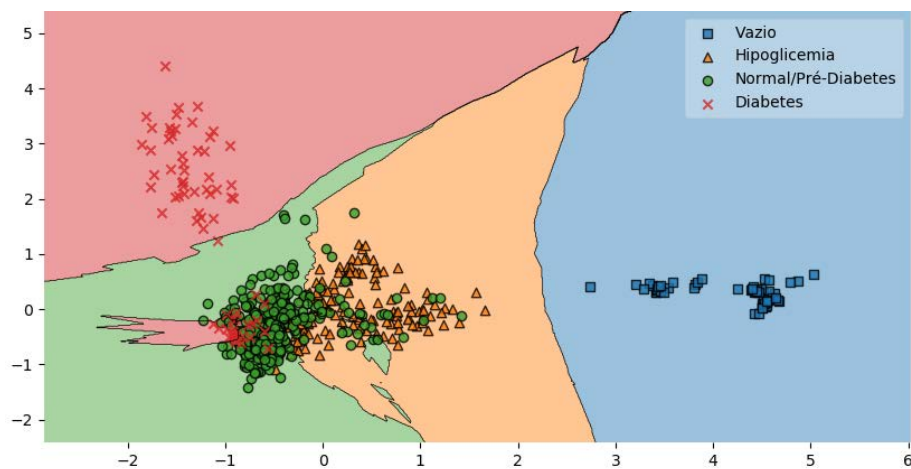


(b) SVM

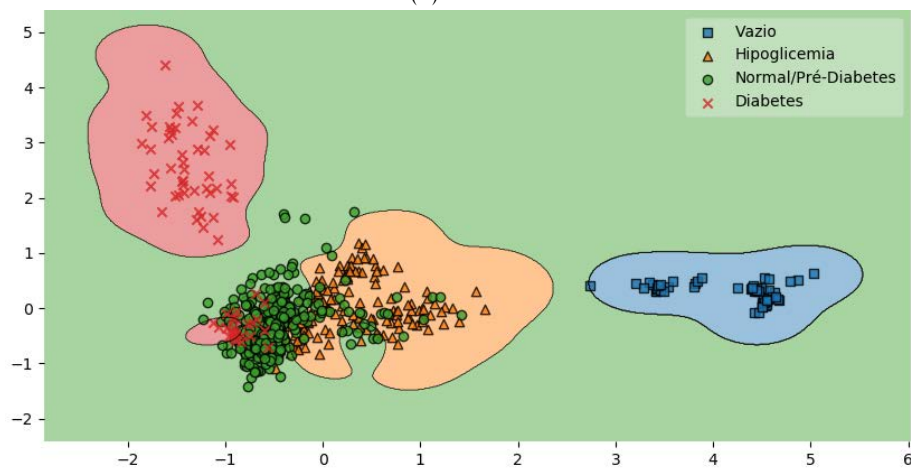


(c) Random Forest

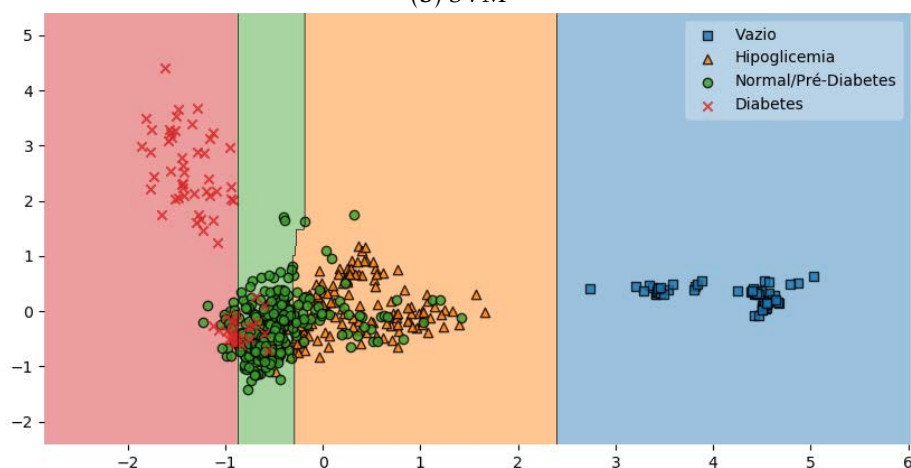
Figura F.33: Regiões de decisão para os quatro níveis de referência da glicose (duas horas após a refeição) dos três algoritmos estudados, baseados na média dos componentes XYZ.



(a) KNN

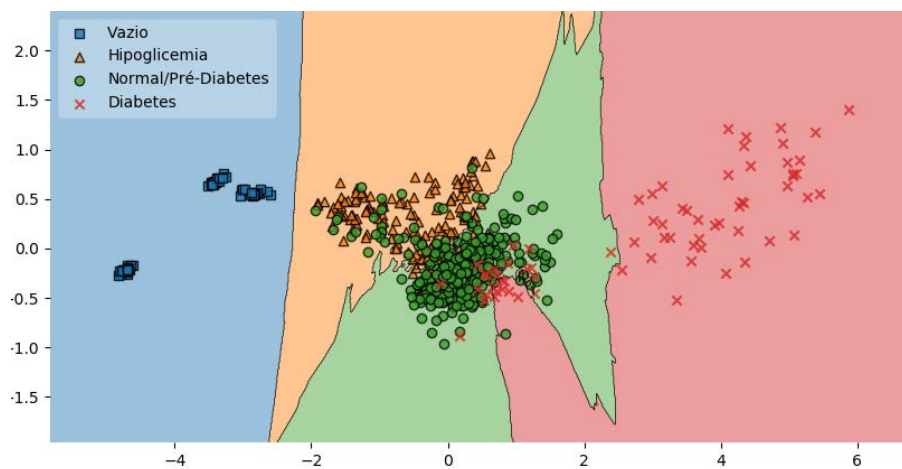


(b) SVM

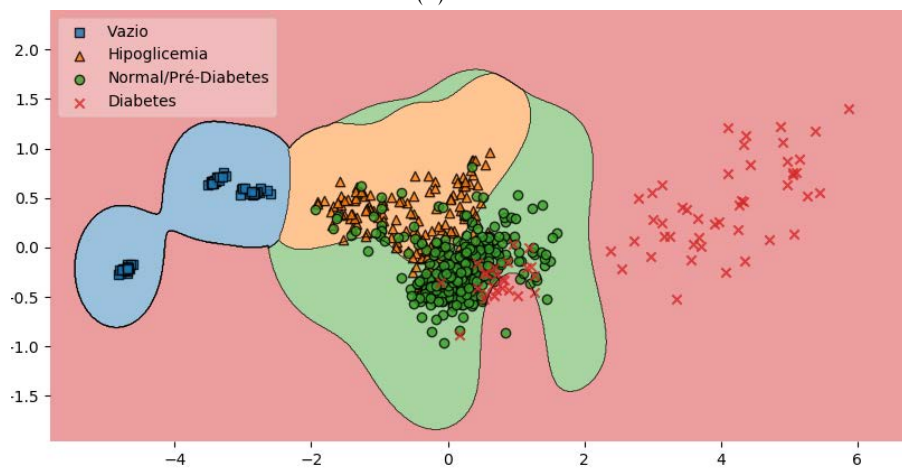


(c) Random Forest

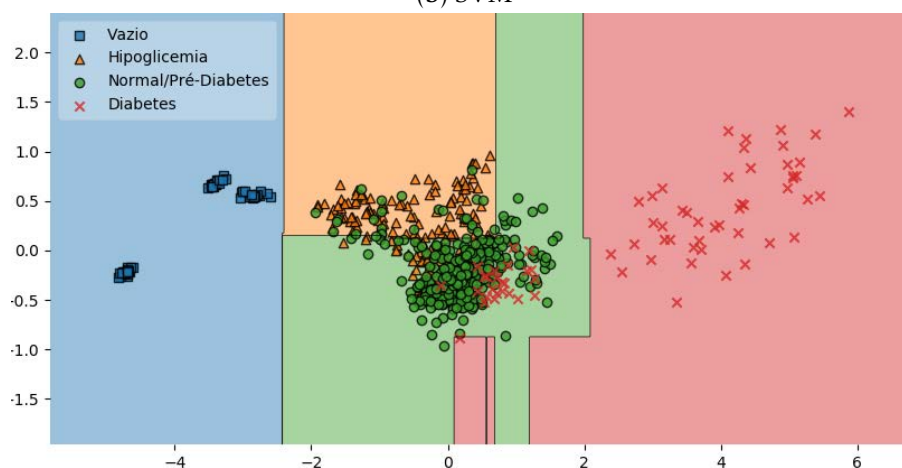
Figura F.34: Regiões de decisão para os quatro níveis de referência da glicose (duas horas após a refeição) dos três algoritmos estudados, baseados na média dos componentes HLS.



(a) KNN



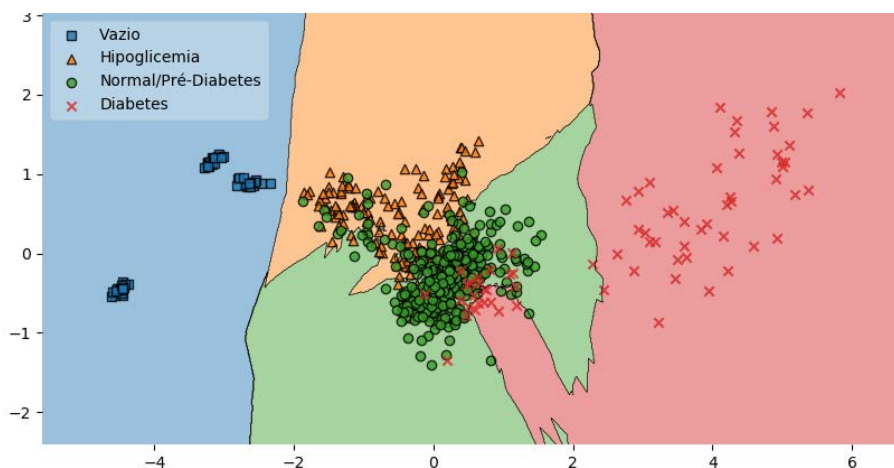
(b) SVM



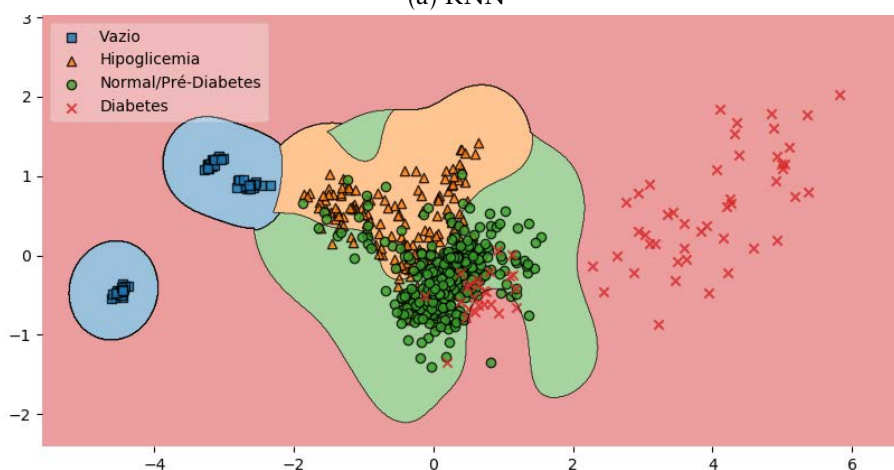
(c) Random Forest

Figura F.35: Regiões de decisão para os quatro níveis de referência da glicose (duas horas após a refeição) dos três algoritmos estudados, baseados na média dos componentes LAB.

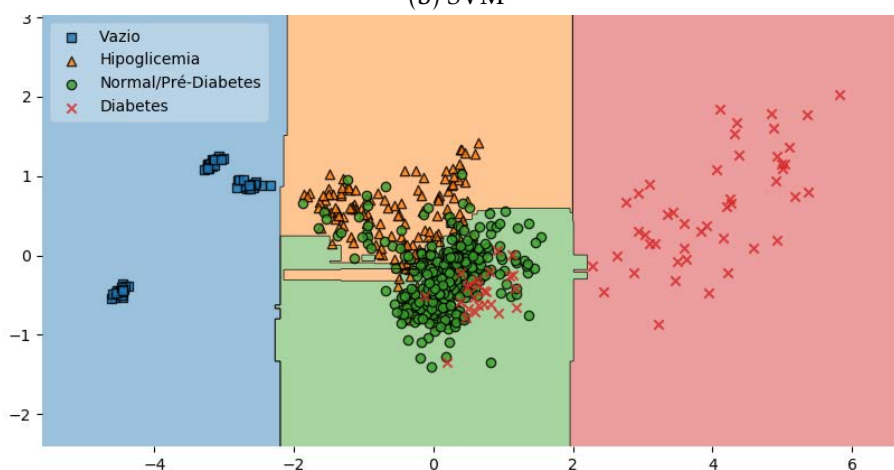




(a) KNN

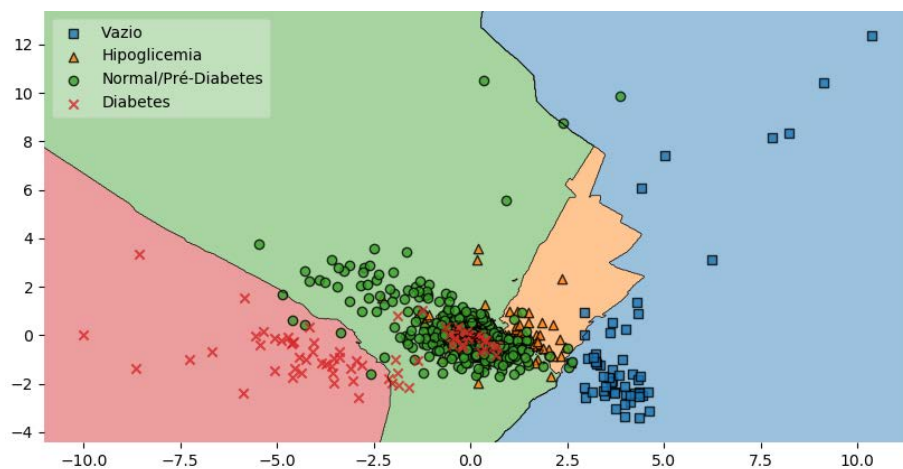


(b) SVM

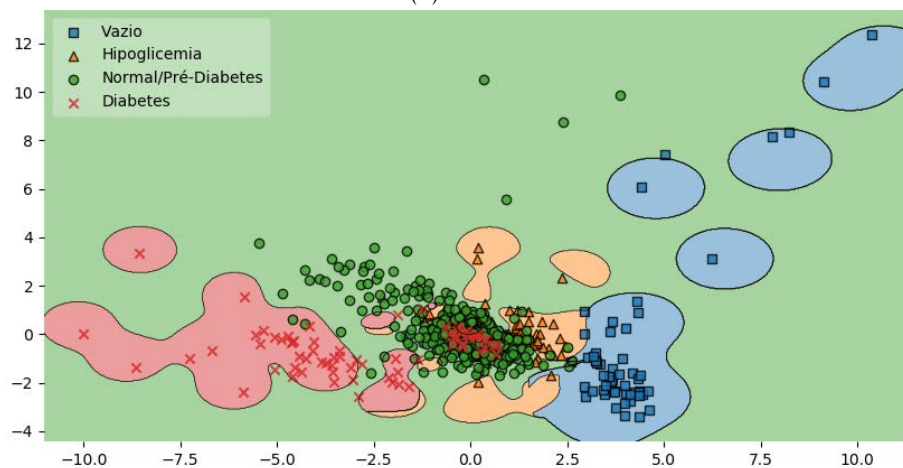


(c) Random Forest

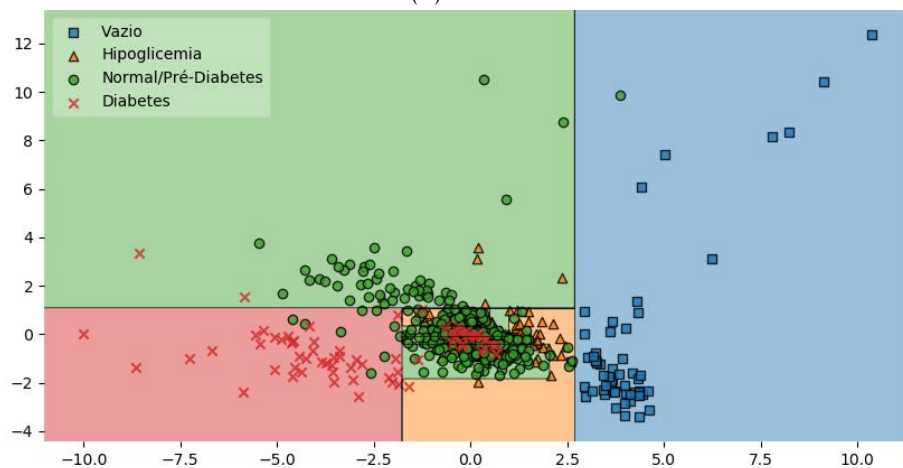
Figura F.36: Regiões de decisão para os quatro níveis de referência da glucose (duas horas após a refeição) dos três algoritmos estudados, baseados na média dos componentes YUV.



(a) KNN



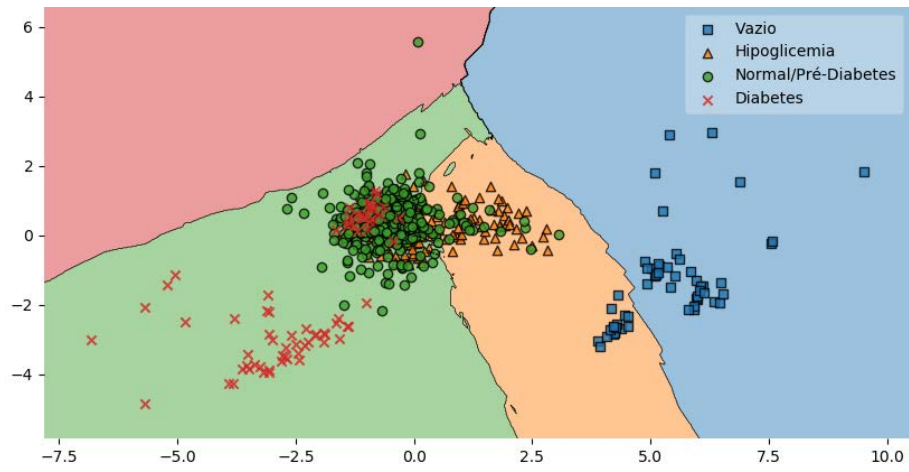
(b) SVM



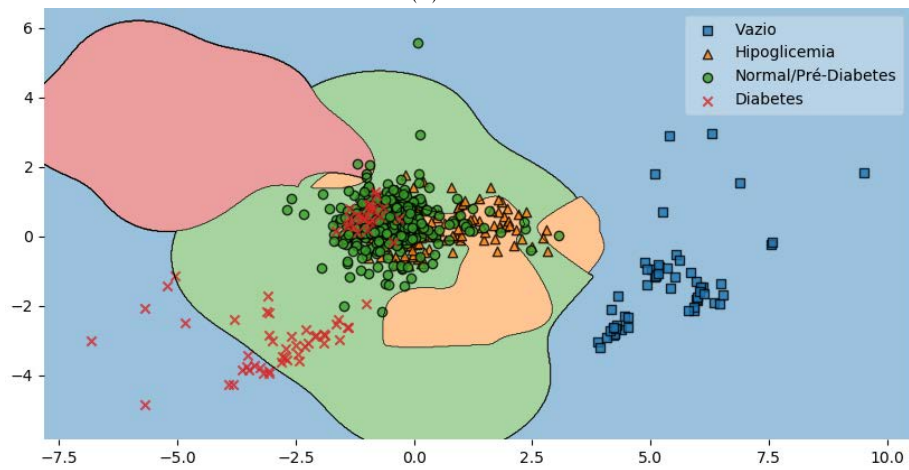
(c) Random Forest

Figura F.37: Regiões de decisão para os quatro níveis de referência da glicose (duas horas após a refeição) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes RGB.

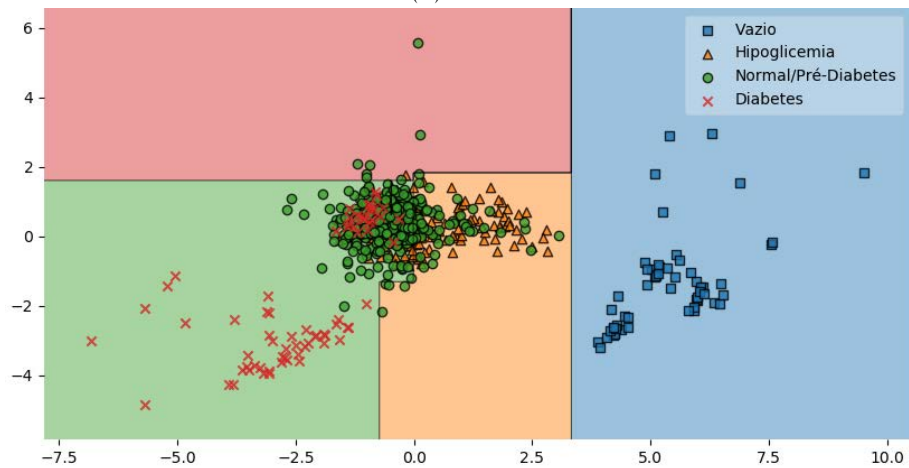




(a) KNN

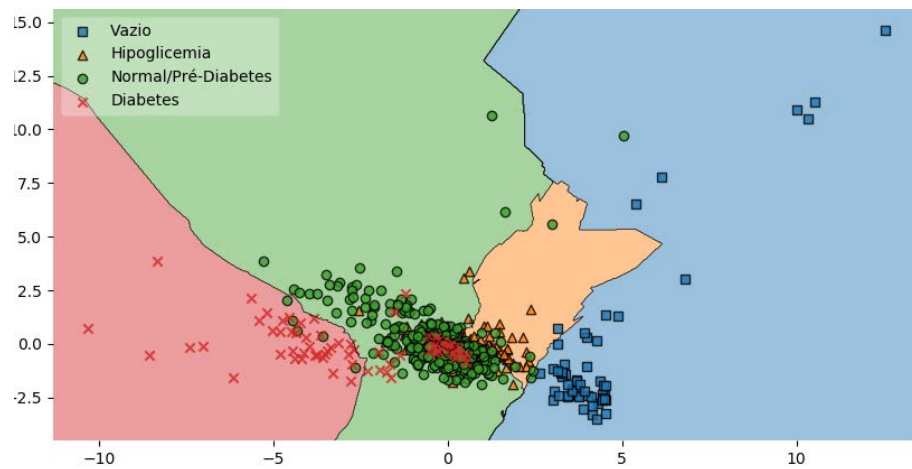


(b) SVM

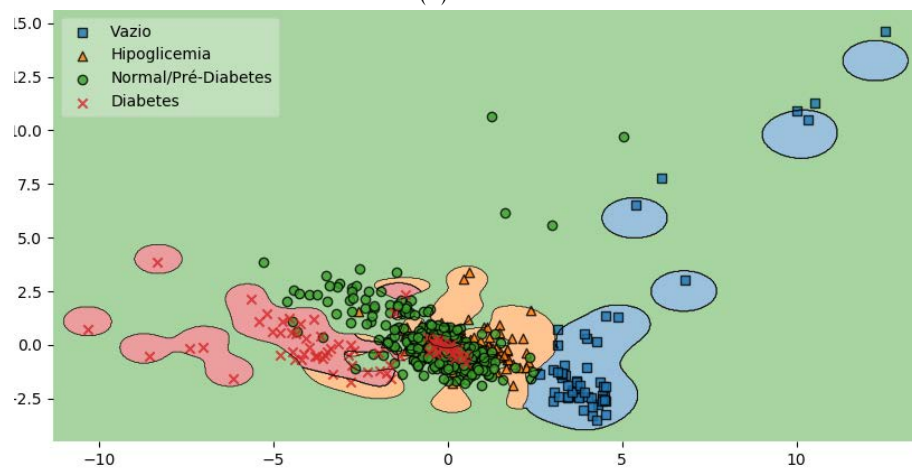


(c) Random Forest

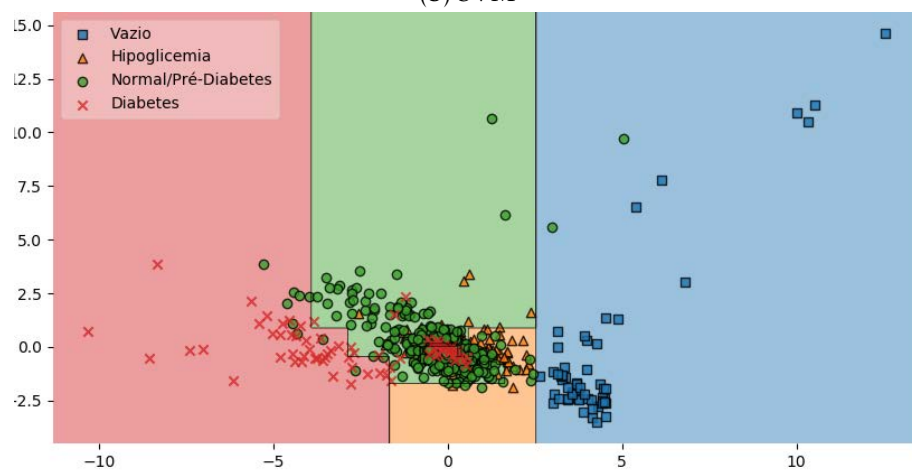
Figura F.38: Regiões de decisão para os quatro níveis de referência da glucose (duas horas após a refeição) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes HSV.



(a) KNN

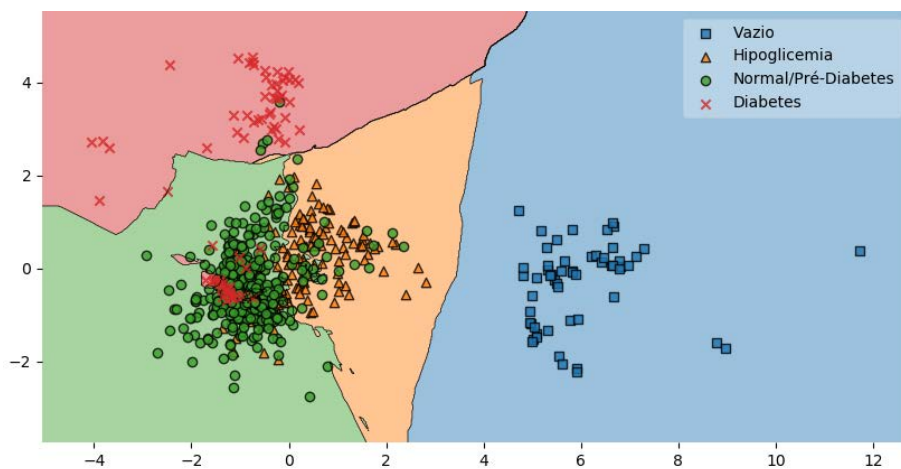


(b) SVM

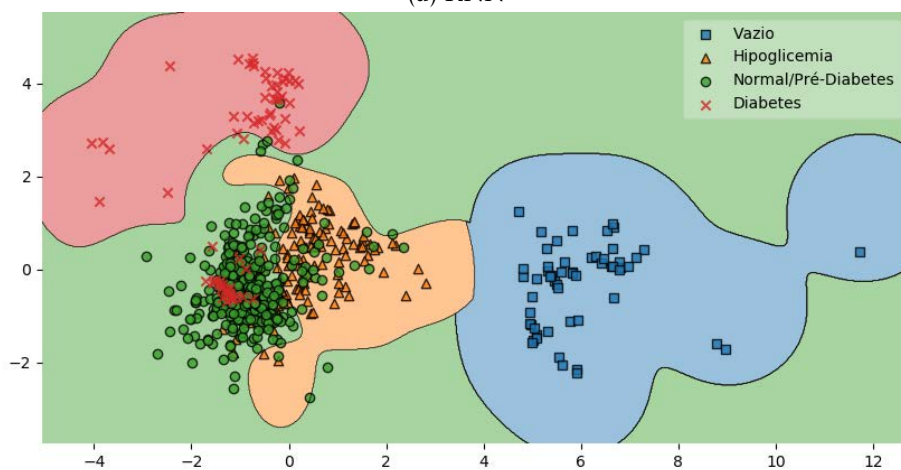


(c) Random Forest

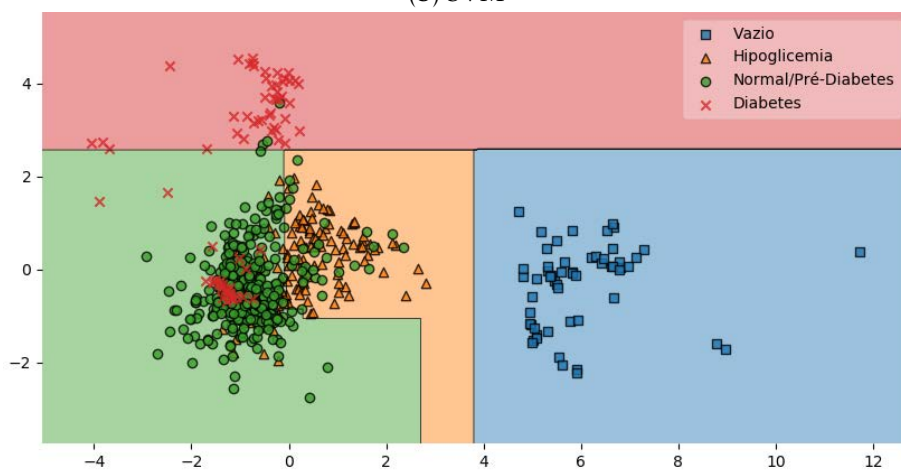
Figura F.39: Regiões de decisão para os quatro níveis de referência da glucose (duas horas após a refeição) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes XYZ.



(a) KNN

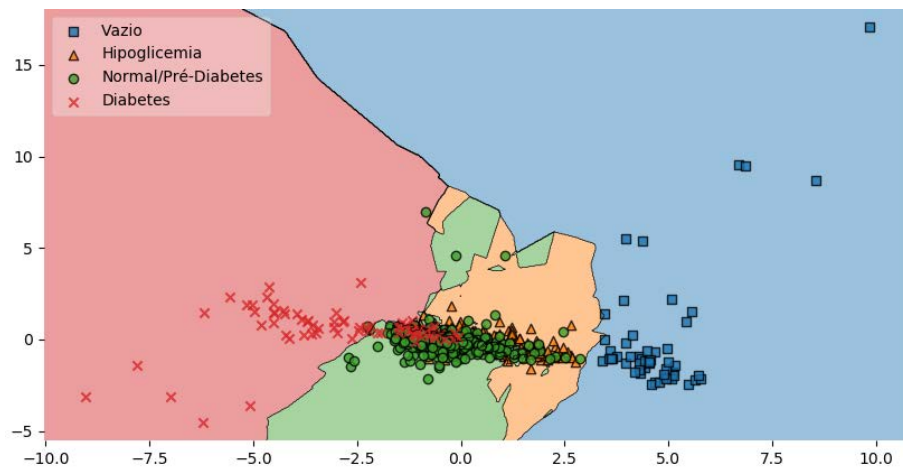


(b) SVM

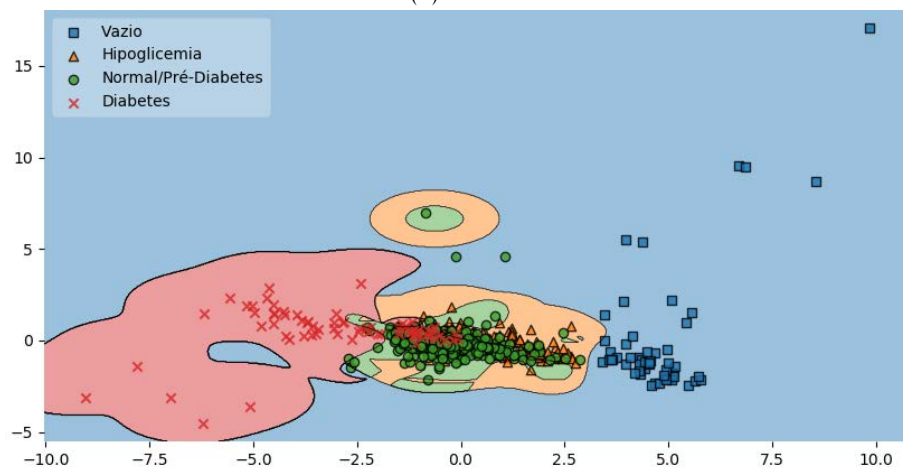


(c) Random Forest

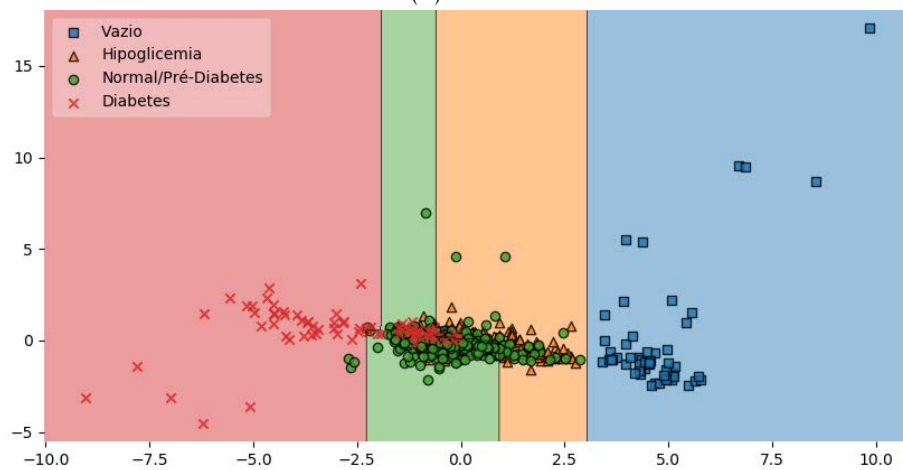
Figura F.40: Regiões de decisão para os quatro níveis de referência da glucose (duas horas após a refeição) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes HLS.



(a) KNN

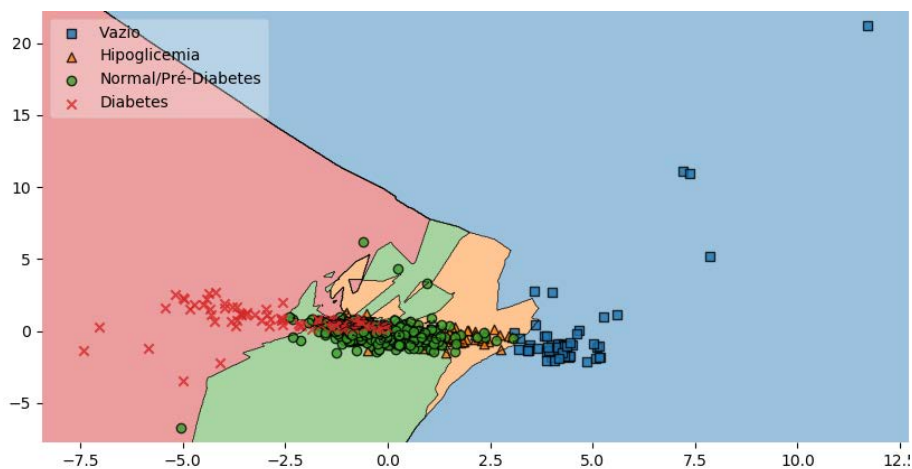


(b) SVM

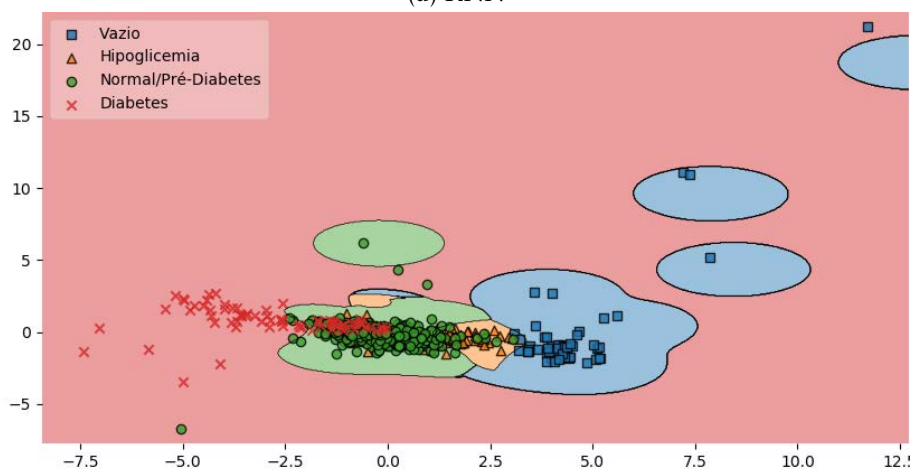


(c) *Random Forest*

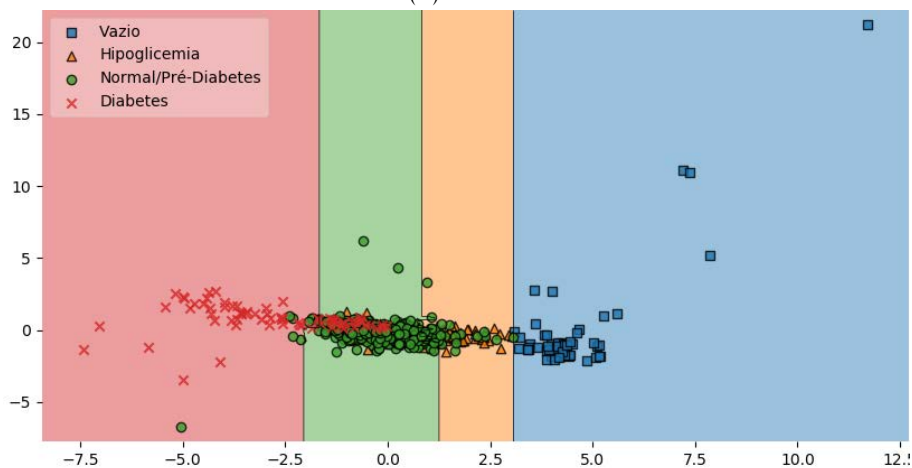
Figura F.41: Regiões de decisão para os quatro níveis de referência da glucose (duas horas após a refeição) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes LAB.



(a) KNN

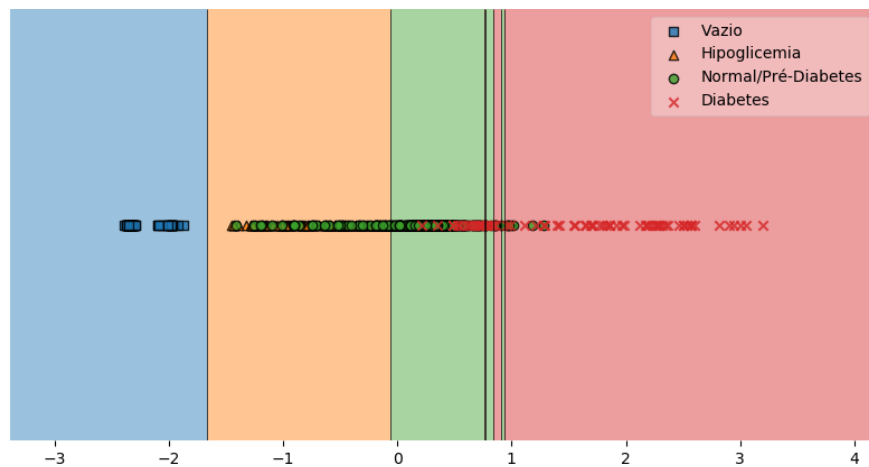


(b) SVM

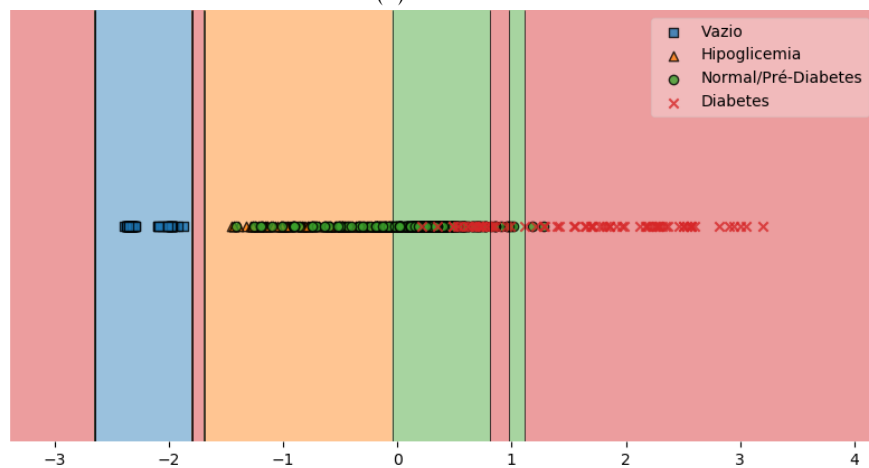


(c) Random Forest

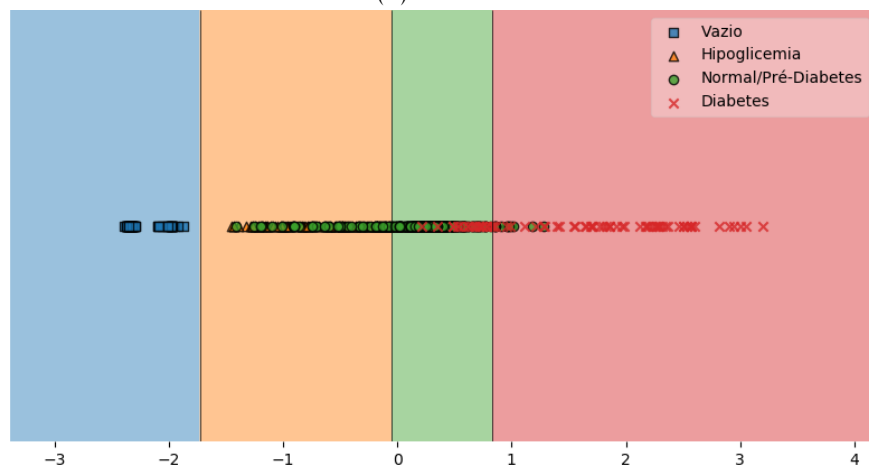
Figura F.42: Regiões de decisão para os quatro níveis de referência da glucose (duas horas após a refeição) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes YUV.



(a) KNN



(b) SVM



(c) Random Forest

Figura F.43: Regiões de decisão para os quatro níveis de referência da glicose (duas horas após a refeição) dos três algoritmos estudados, baseados no rácio entre os componentes vermelho e verde do espaço RGB.



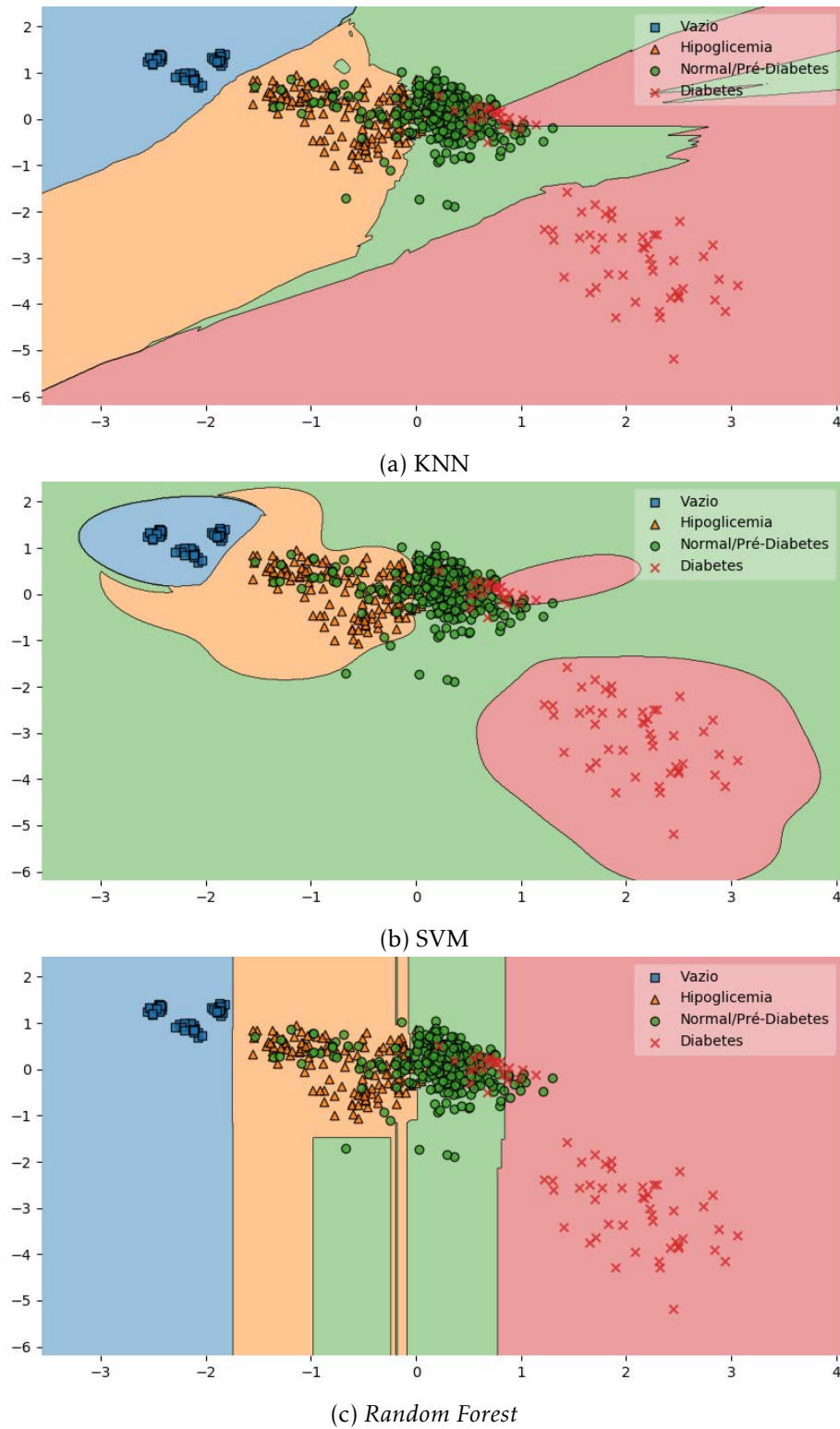
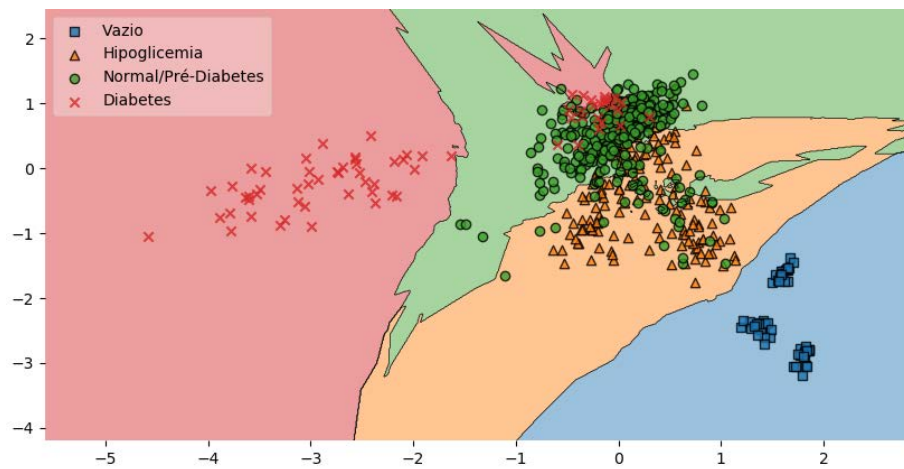
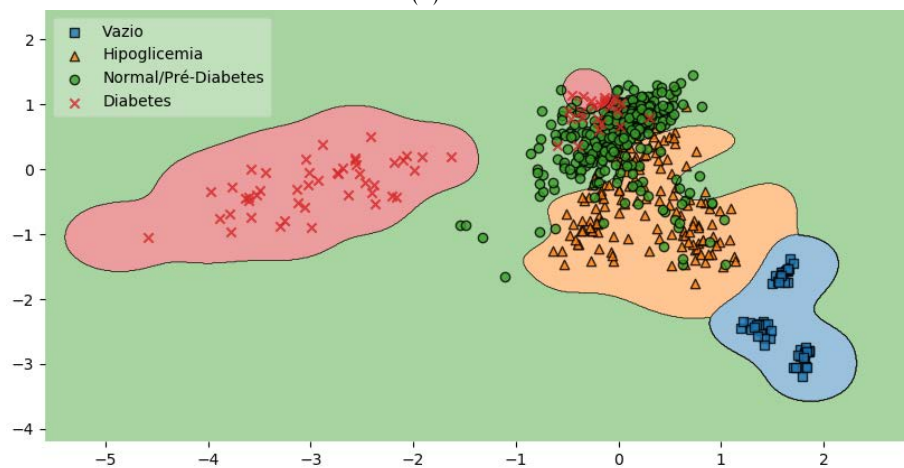


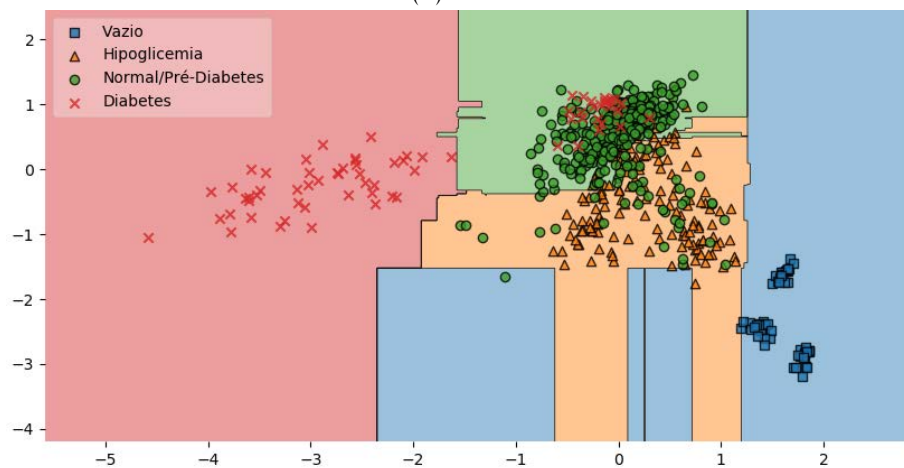
Figura F.44: Regiões de decisão para os quatro níveis de referência da glucose (duas horas após a refeição) dos três algoritmos estudados, baseados nos componentes S e V do espaço HSV.



(a) KNN



(b) SVM



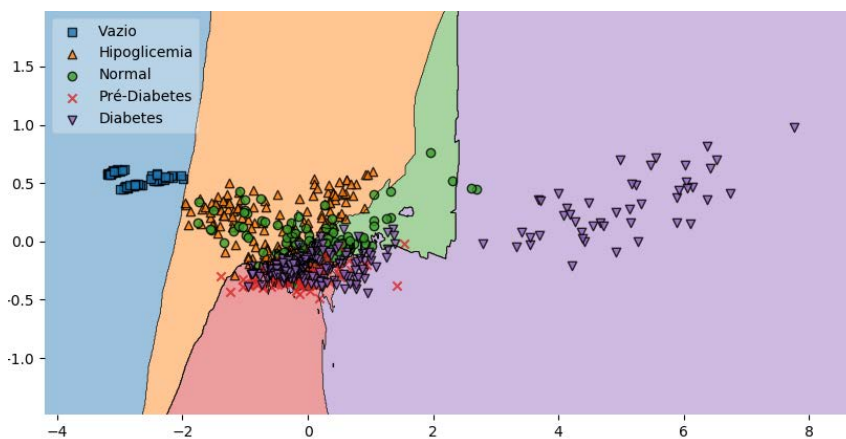
(c) Random Forest

Figura F.45: Regiões de decisão para os quatro níveis de referência da glicose (duas horas após a refeição) dos três algoritmos estudados, baseados nos componentes L e S do espaço HLS.

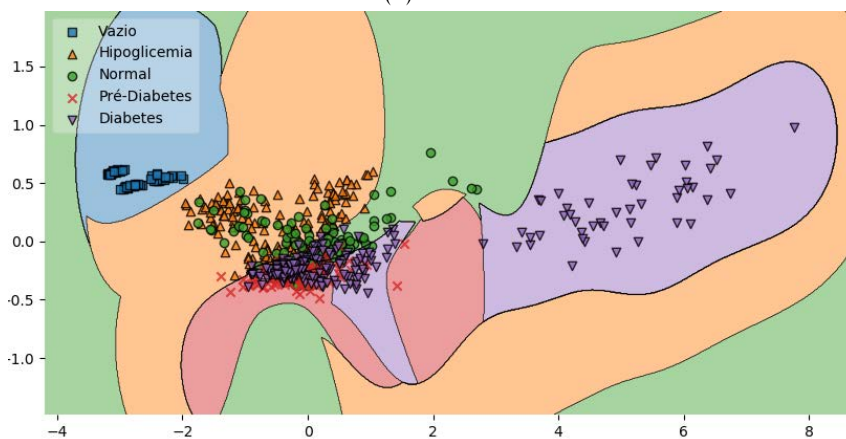


## F.3 Cinco Níveis de Glucose

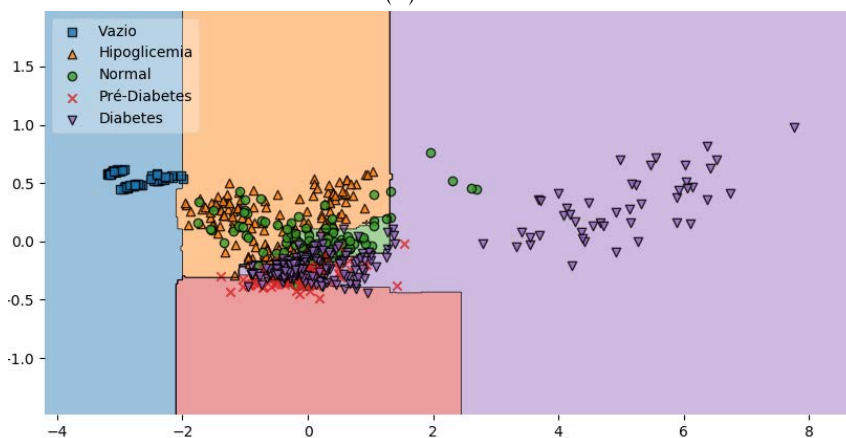
### F.3.1 Estado de jejum



(a) KNN

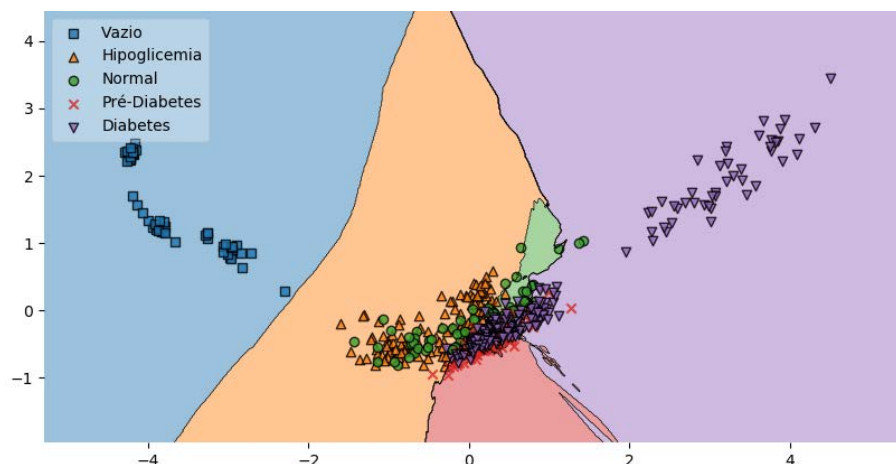


(b) SVM

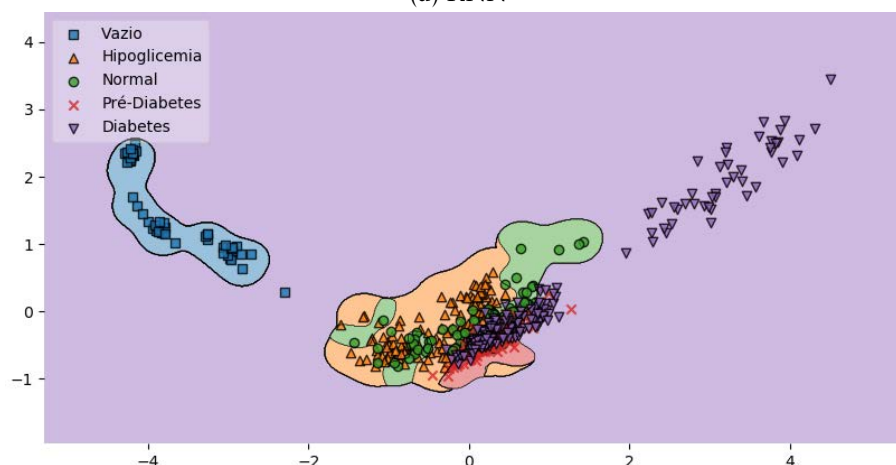


(c) Random Forest

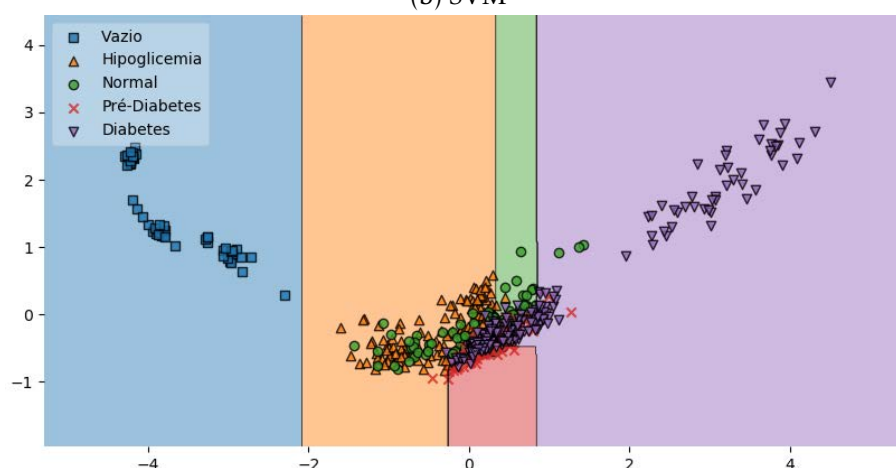
Figura F.46: Regiões de decisão para os cinco níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados na média dos componentes RGB.



(a) KNN

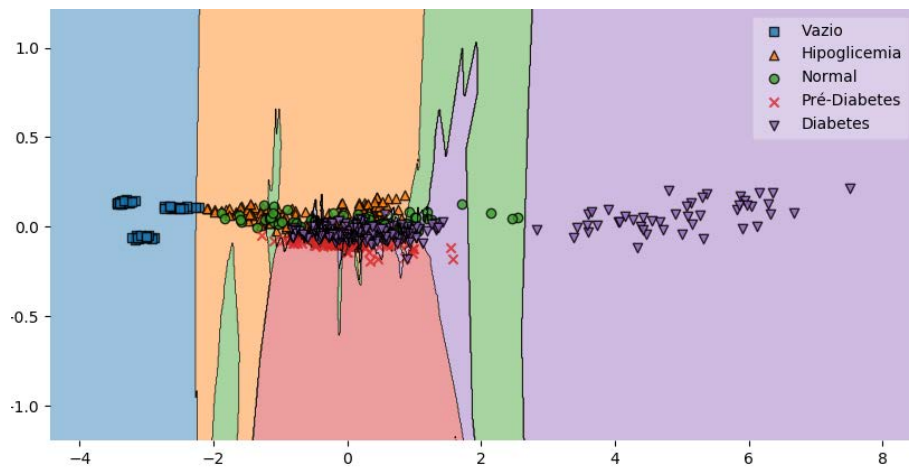


(b) SVM

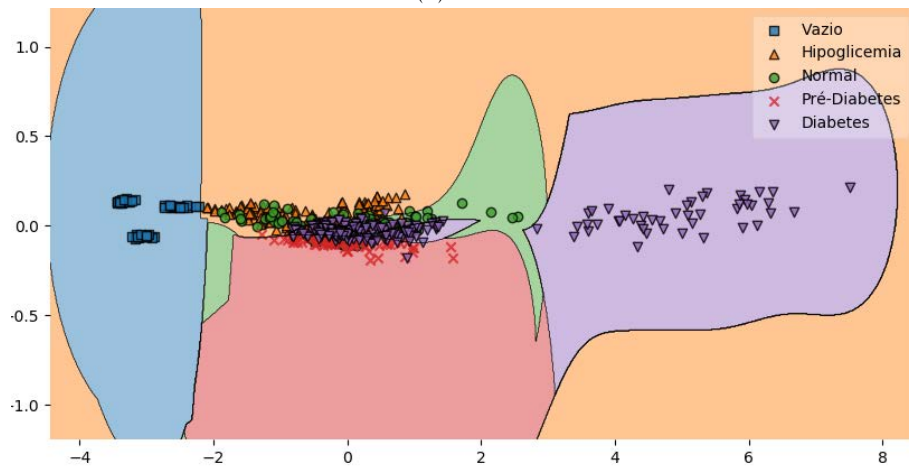


(c) *Random Forest*

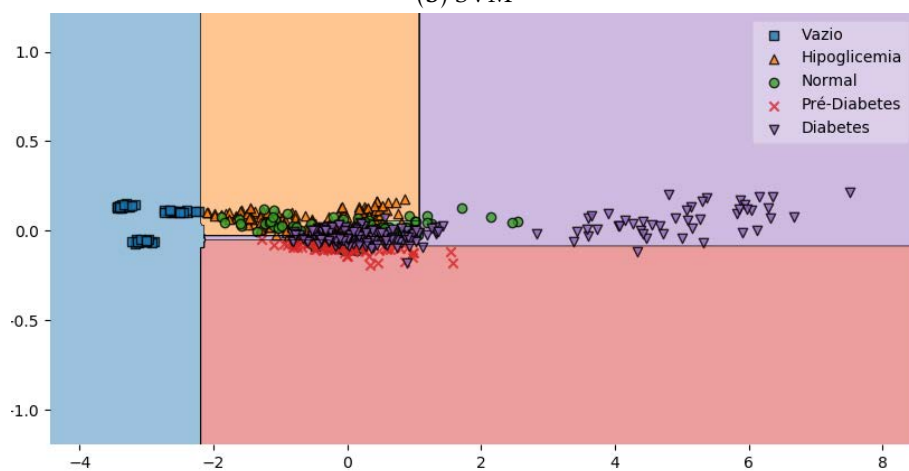
Figura F.47: Regiões de decisão para os cinco níveis de referência da glicose (em jejum) dos três algoritmos estudados, baseados na média dos componentes HSV.



(a) KNN

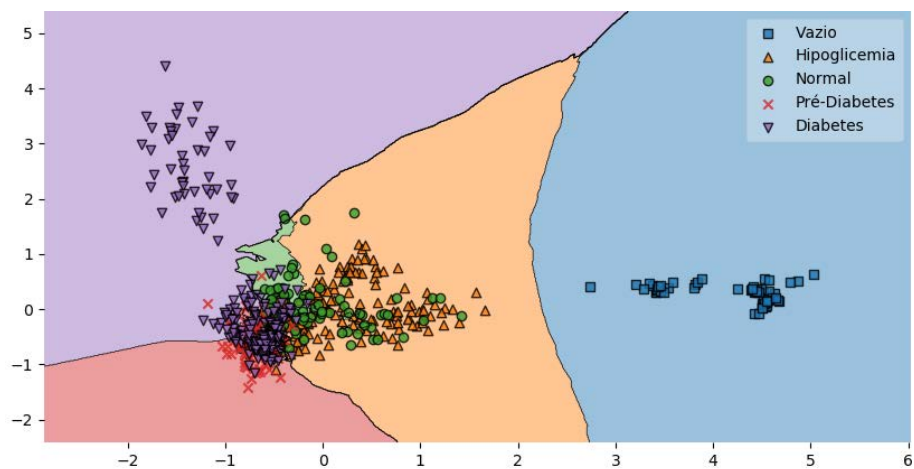


(b) SVM

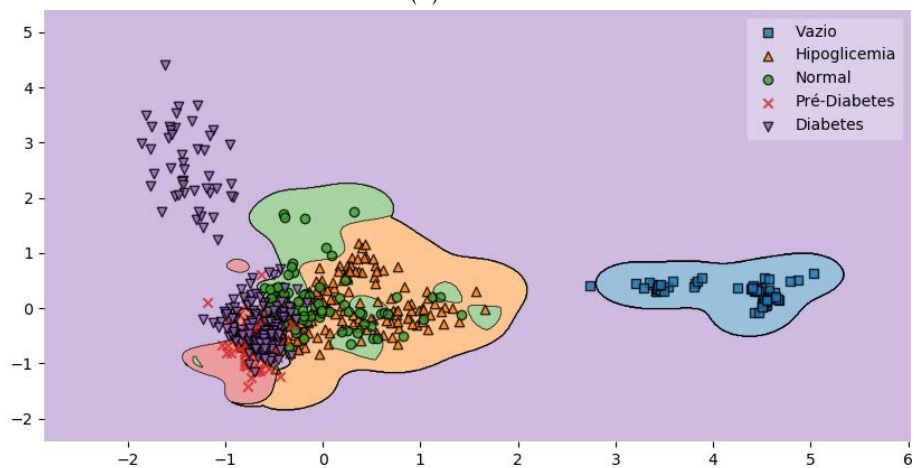


(c) Random Forest

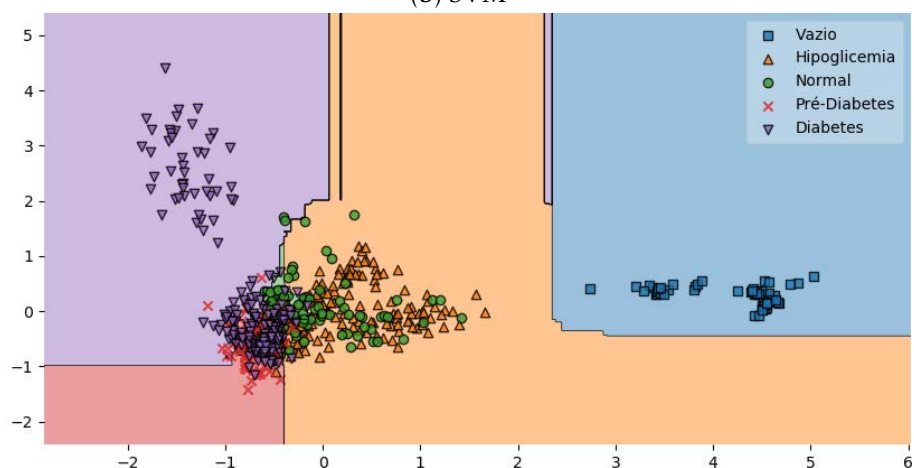
Figura F.48: Regiões de decisão para os cinco níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados na média dos componentes XYZ.



(a) KNN

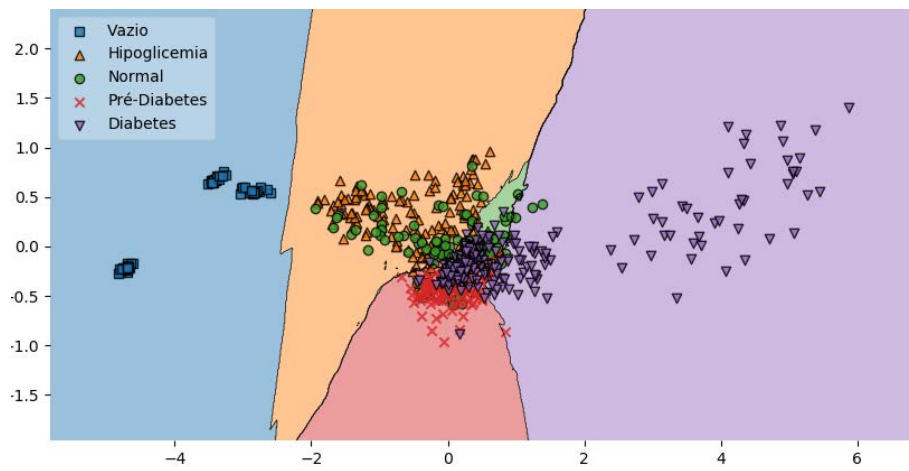


(b) SVM

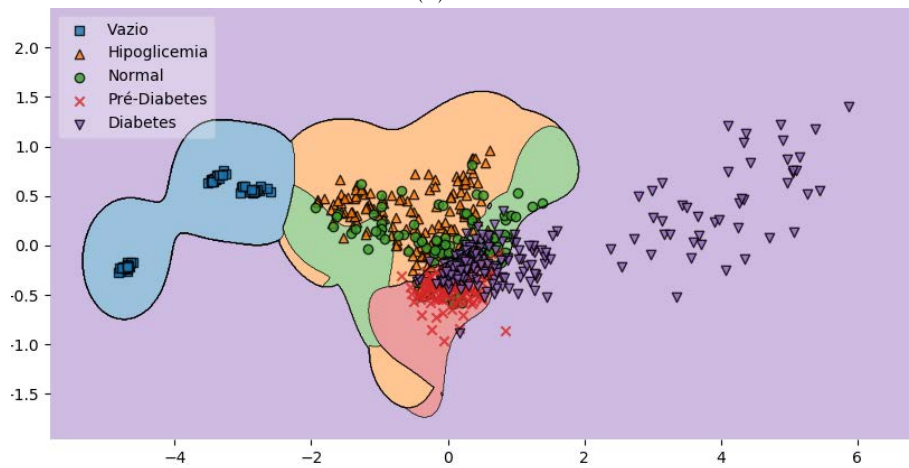


(c) *Random Forest*

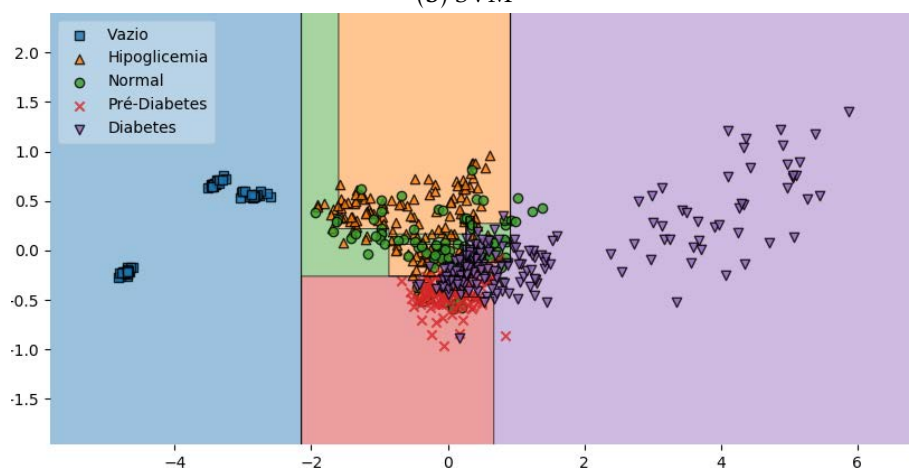
Figura F.49: Regiões de decisão para os cinco níveis de referência da glicose (em jejum) dos três algoritmos estudados, baseados na média dos componentes HLS.



(a) KNN

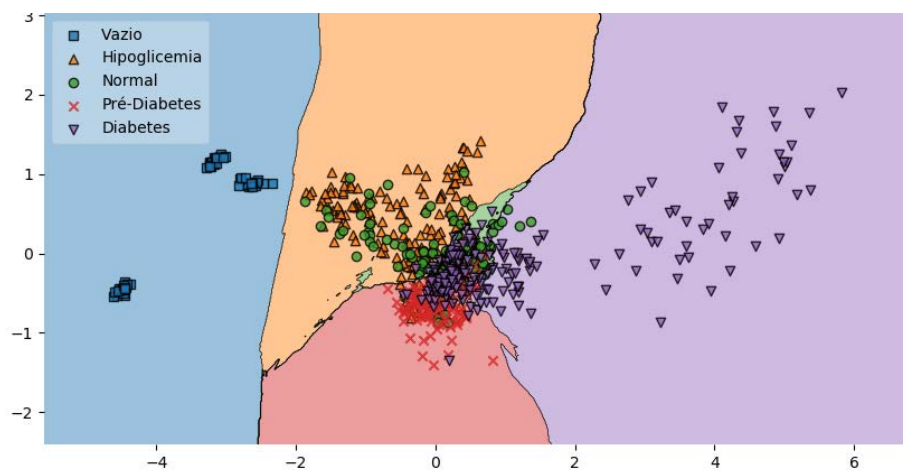


(b) SVM

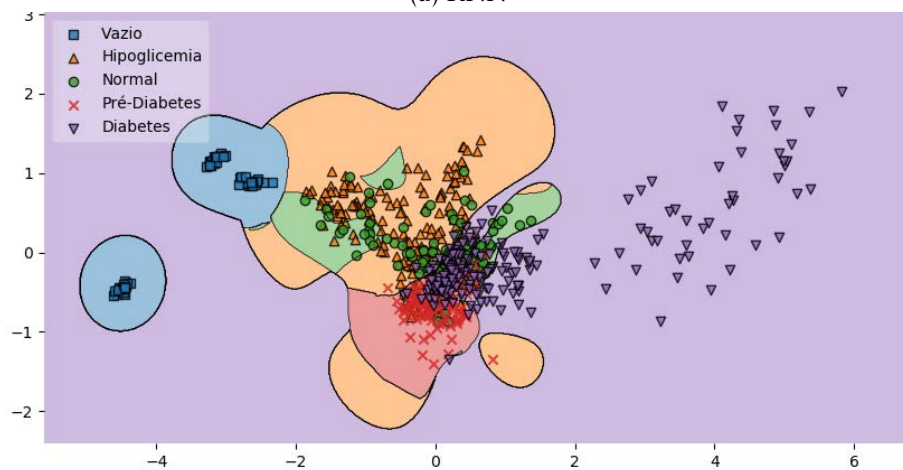


(c) Random Forest

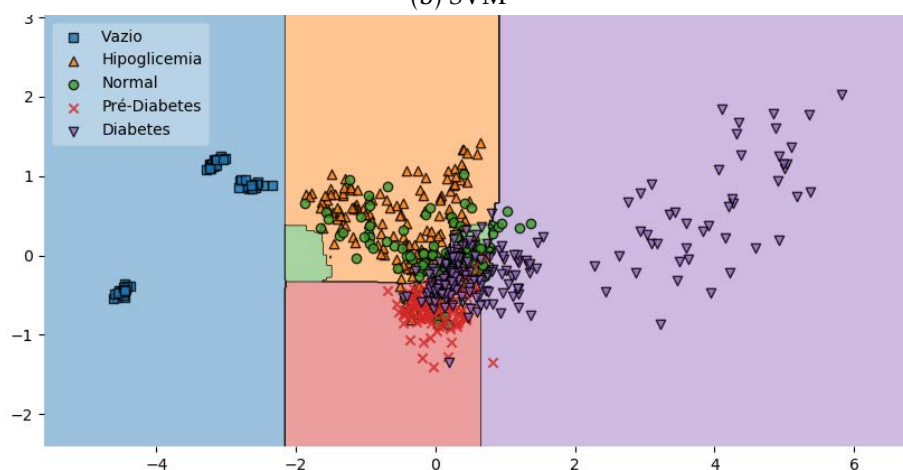
Figura F.50: Regiões de decisão para os cinco níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados na média dos componentes LAB.



(a) KNN



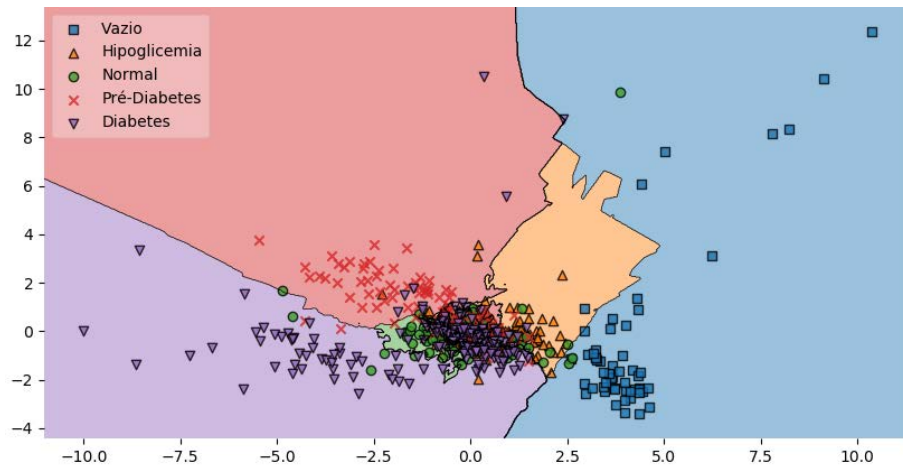
(b) SVM



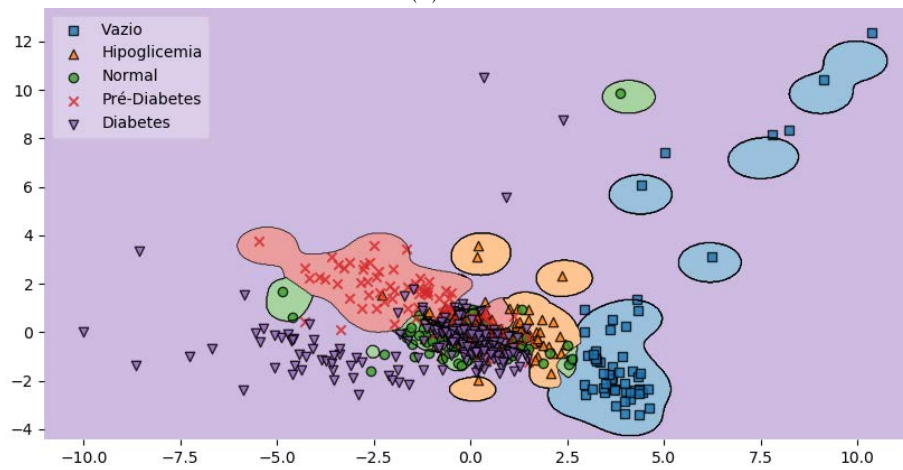
(c) Random Forest

Figura F.51: Regiões de decisão para os cinco níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados na média dos componentes YUV.

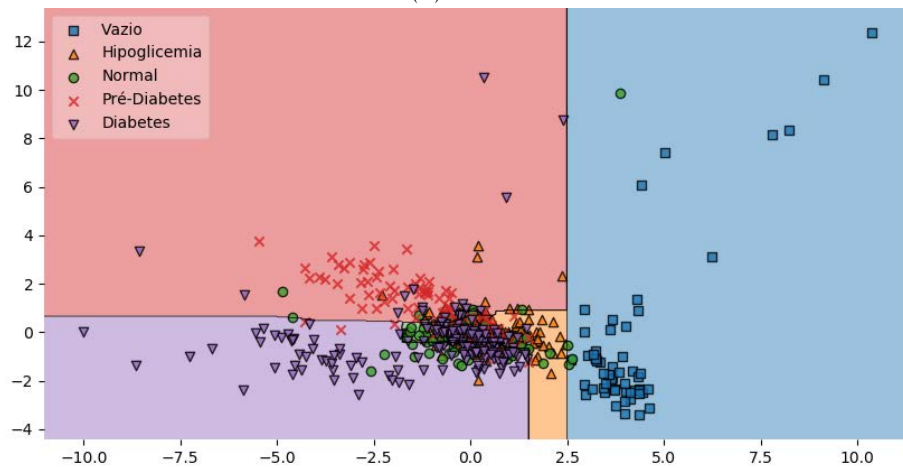




(a) KNN

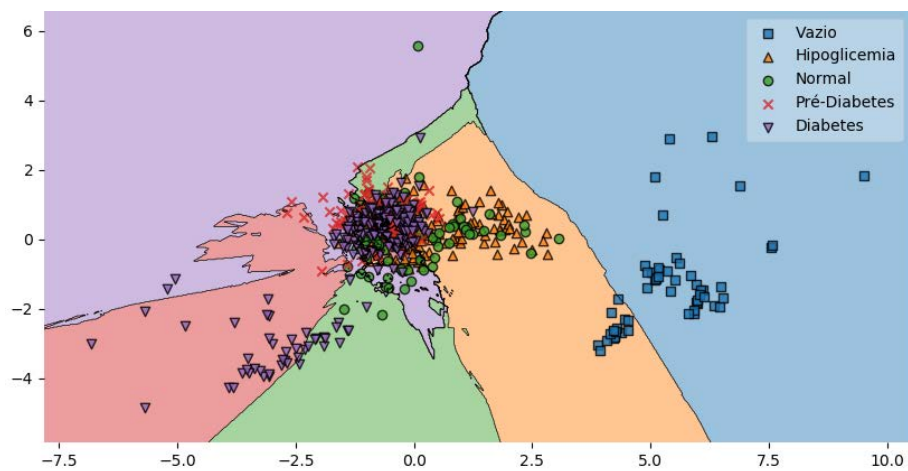


(b) SVM

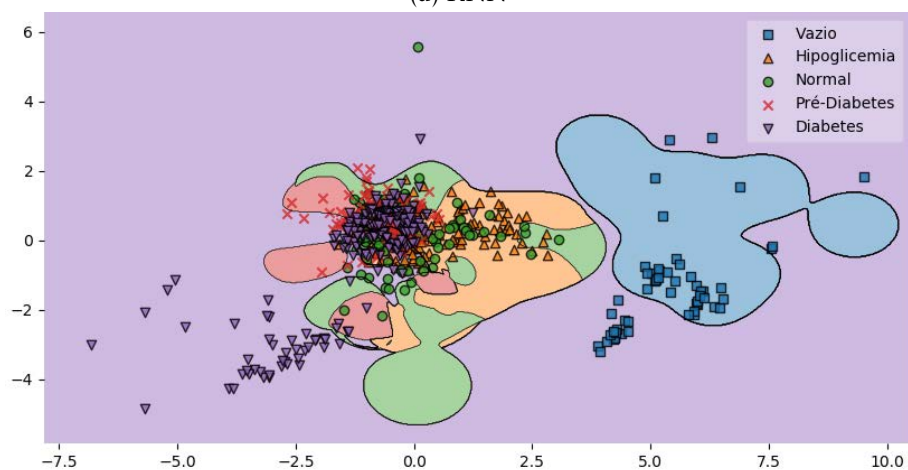


(c) Random Forest

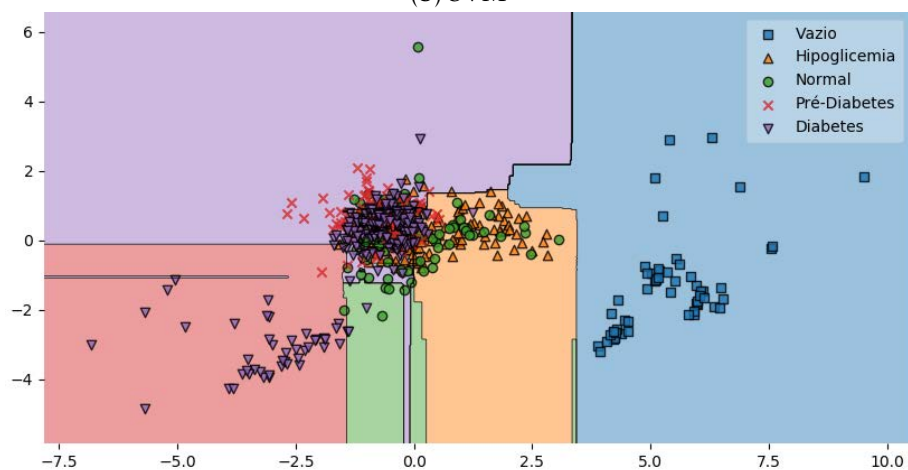
Figura F.52: Regiões de decisão para os cinco níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes RGB.



(a) KNN



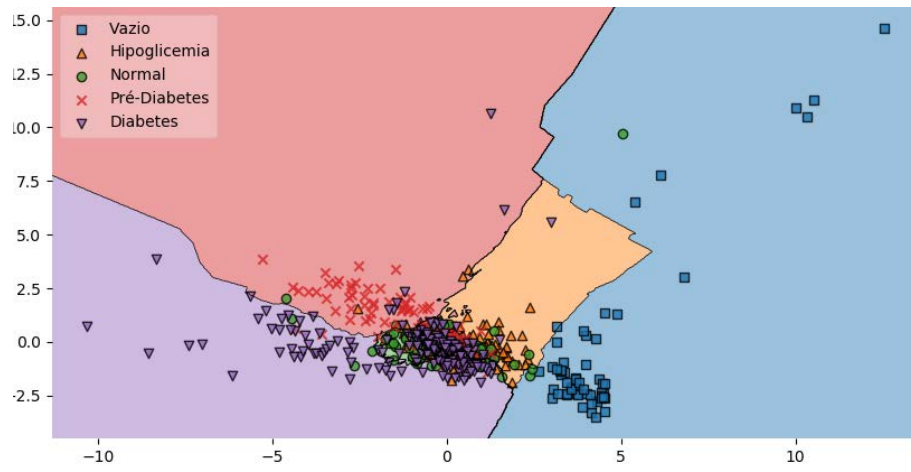
(b) SVM



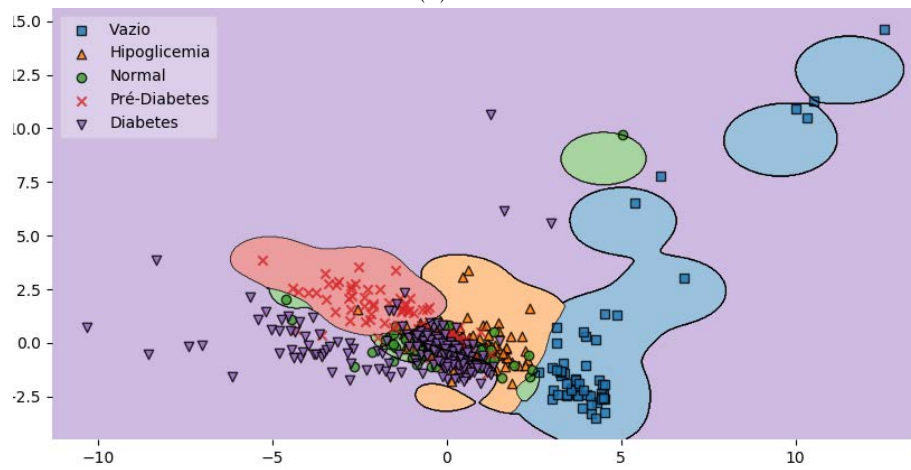
(c) Random Forest

Figura F.53: Regiões de decisão para os cinco níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes HSV.

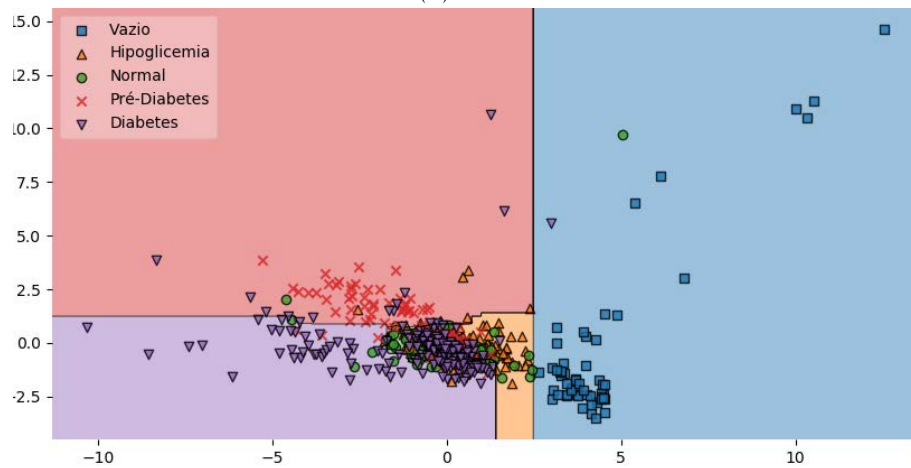




(a) KNN

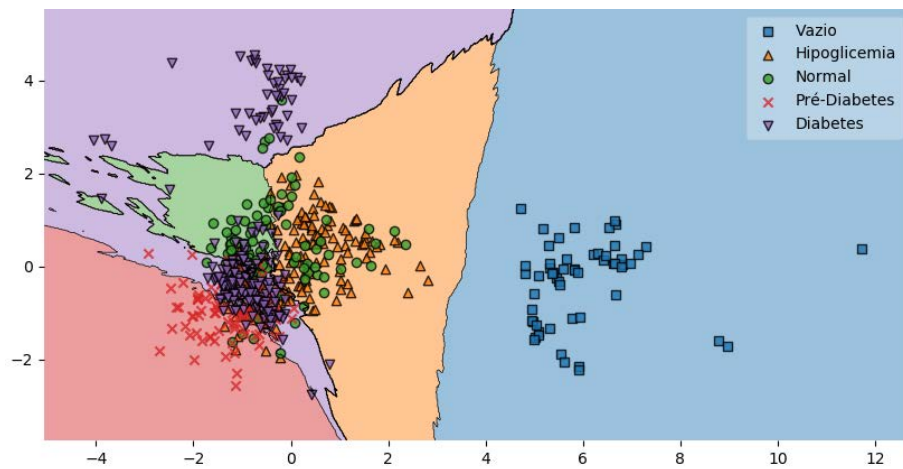


(b) SVM

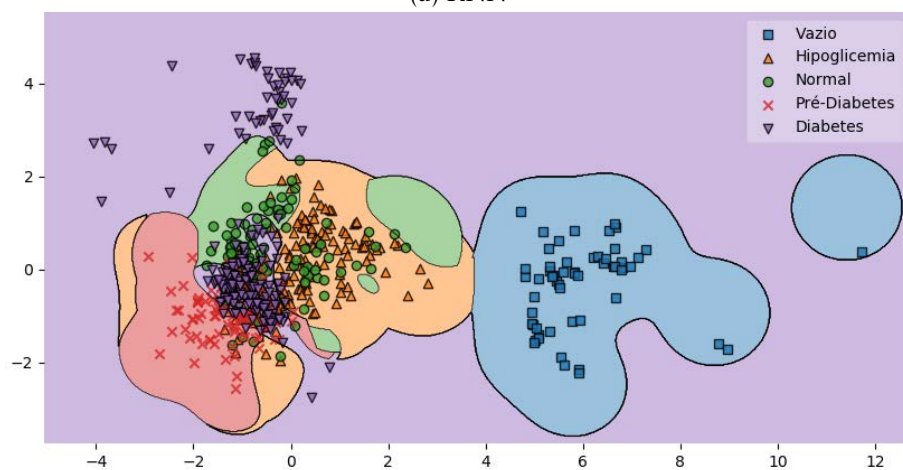


(c) Random Forest

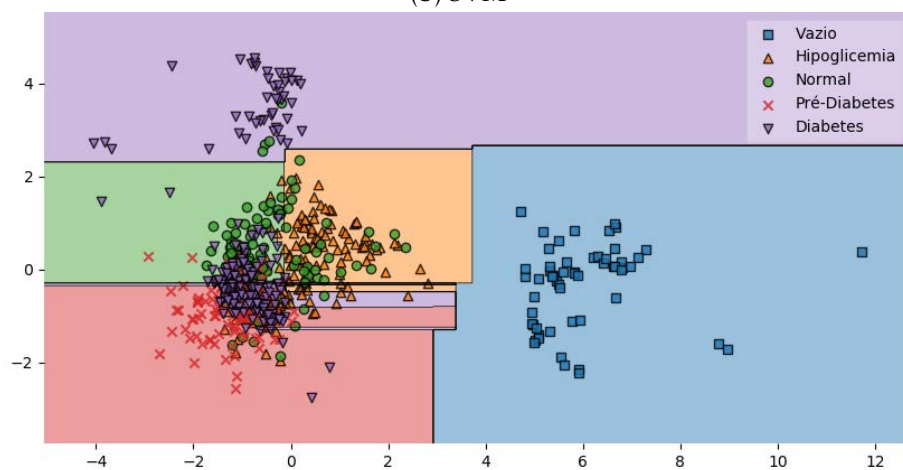
Figura F.54: Regiões de decisão para os cinco níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes XYZ.



(a) KNN

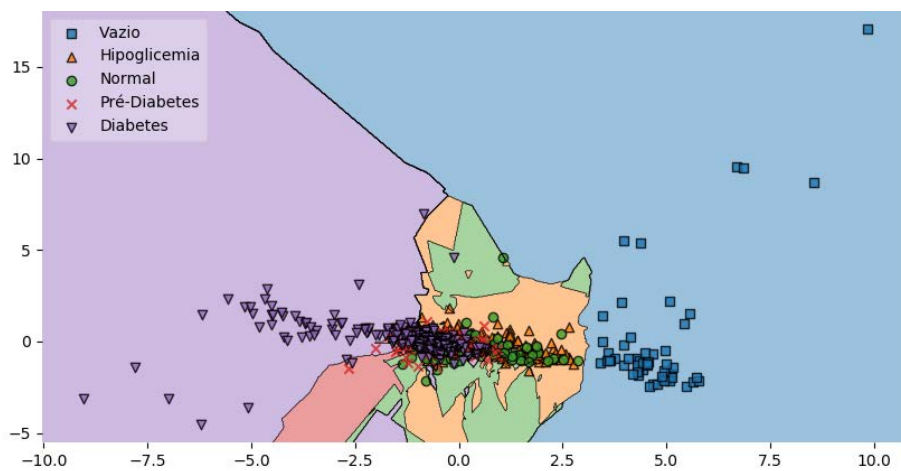


(b) SVM

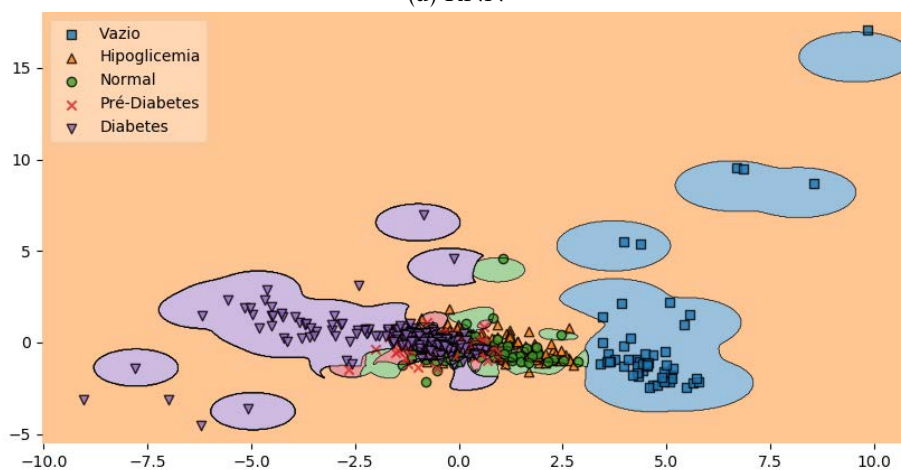


(c) *Random Forest*

Figura F.55: Regiões de decisão para os cinco níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes HLS.



(a) KNN



(b) SVM

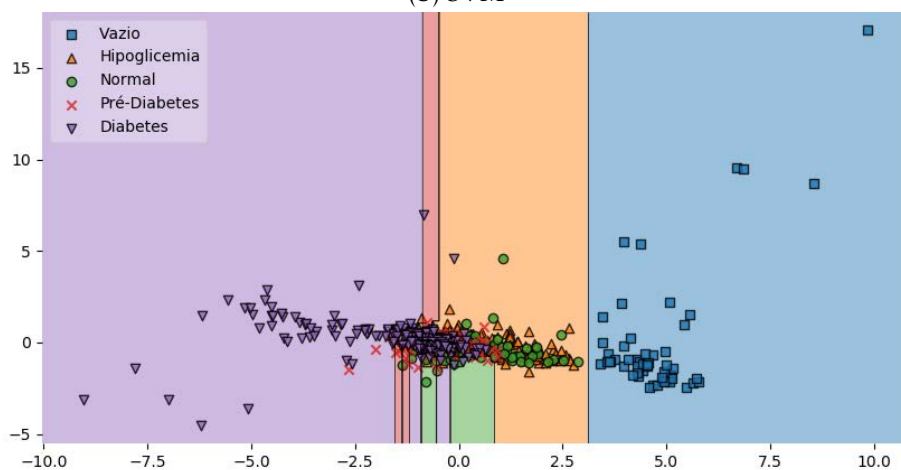
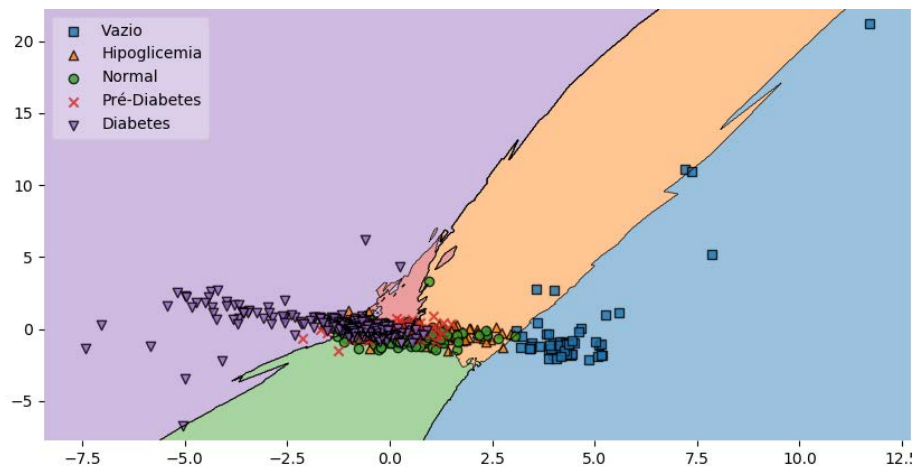
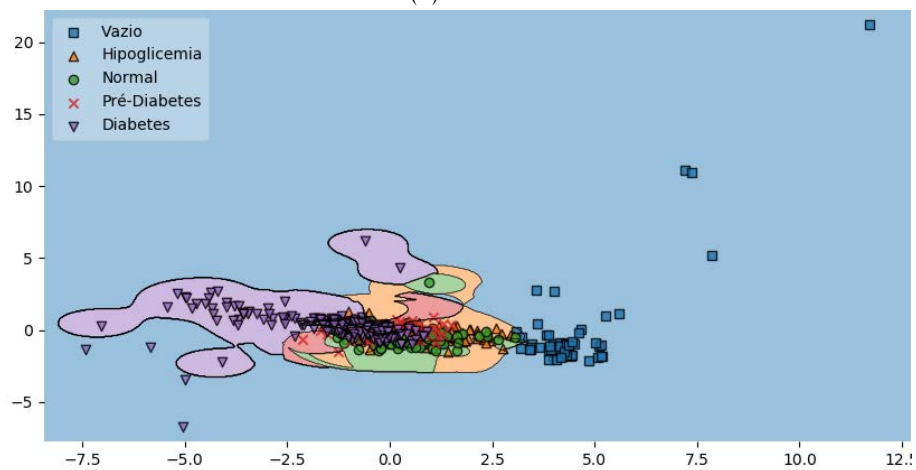
(c) *Random Forest*

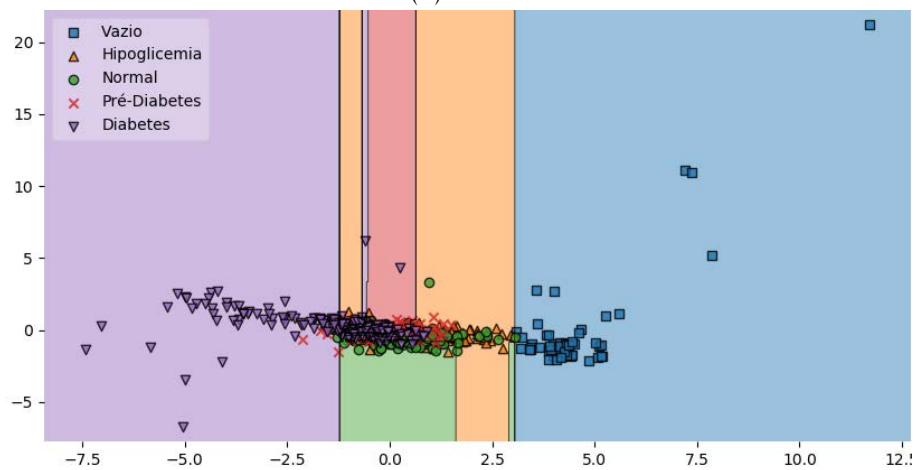
Figura F.56: Regiões de decisão para os cinco níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes LAB.



(a) KNN

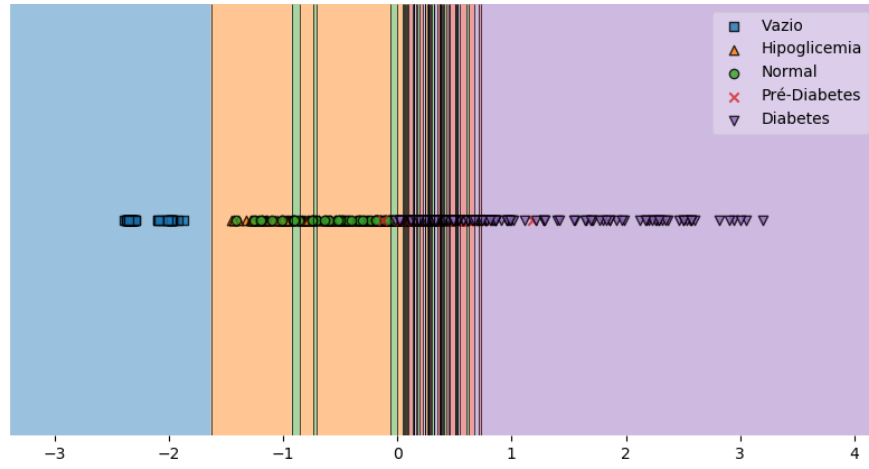


(b) SVM

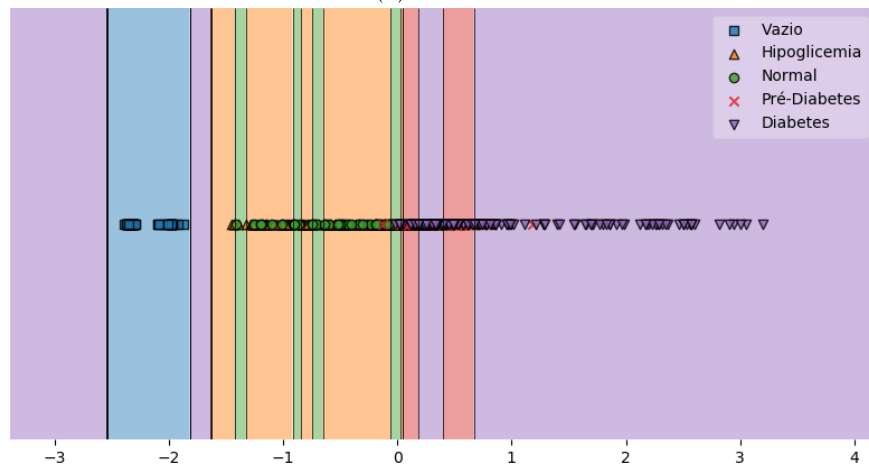


(c) Random Forest

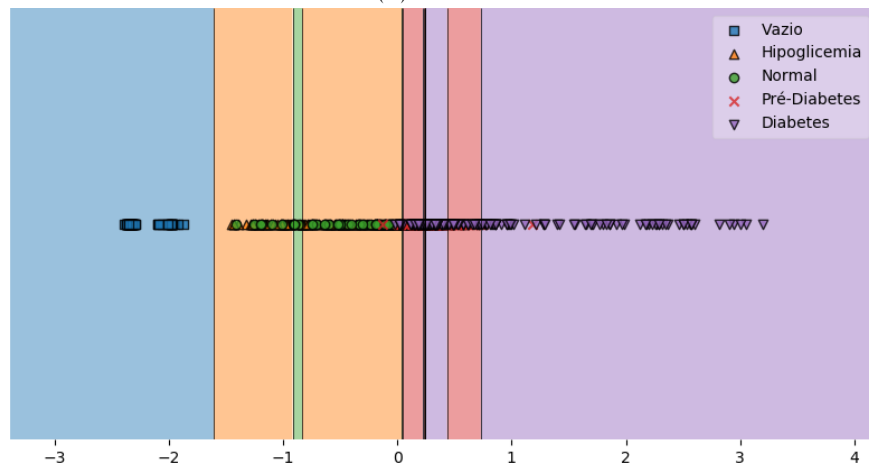
Figura F.57: Regiões de decisão para os cinco níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes YUV.



(a) KNN

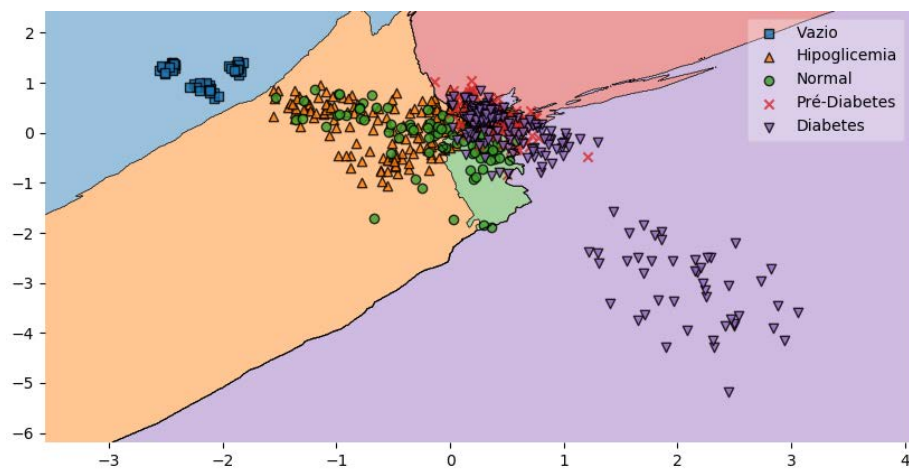


(b) SVM

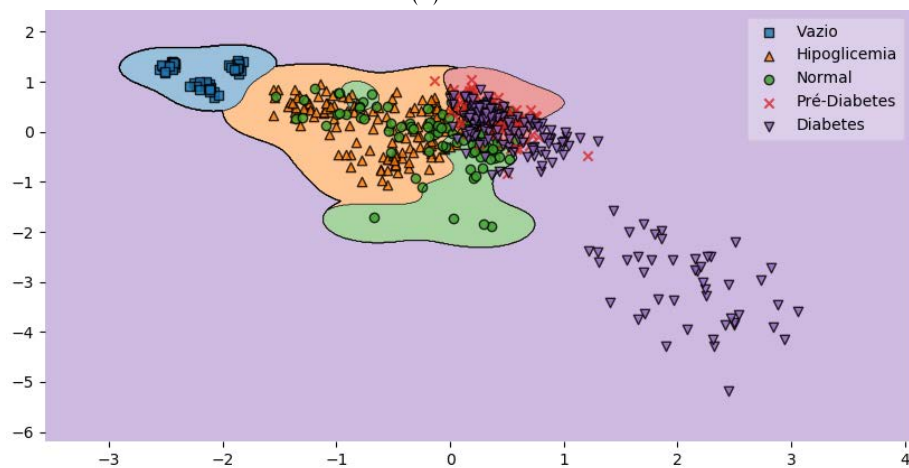


(c) Random Forest

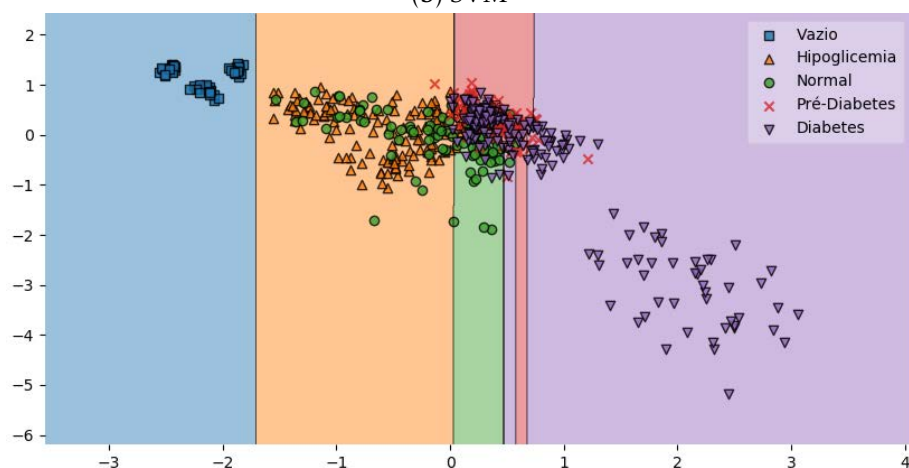
Figura F.58: Regiões de decisão para os cinco níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados no rácio entre os componentes vermelho e verde do espaço RGB.



(a) KNN



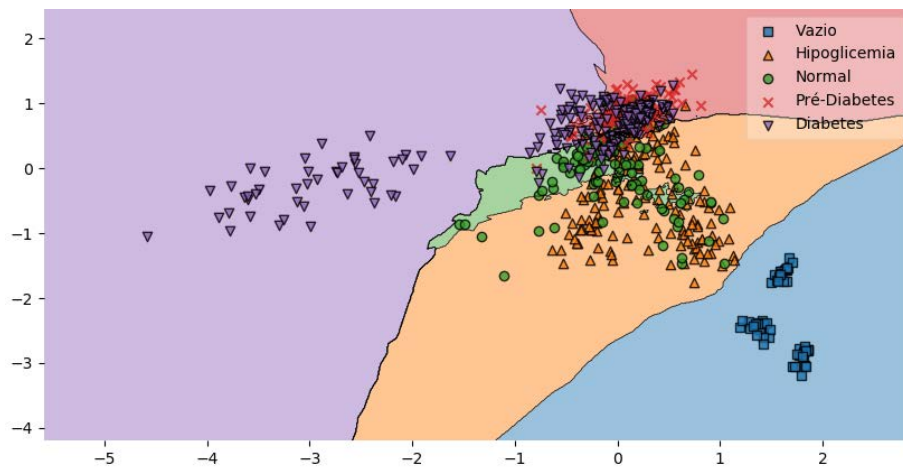
(b) SVM



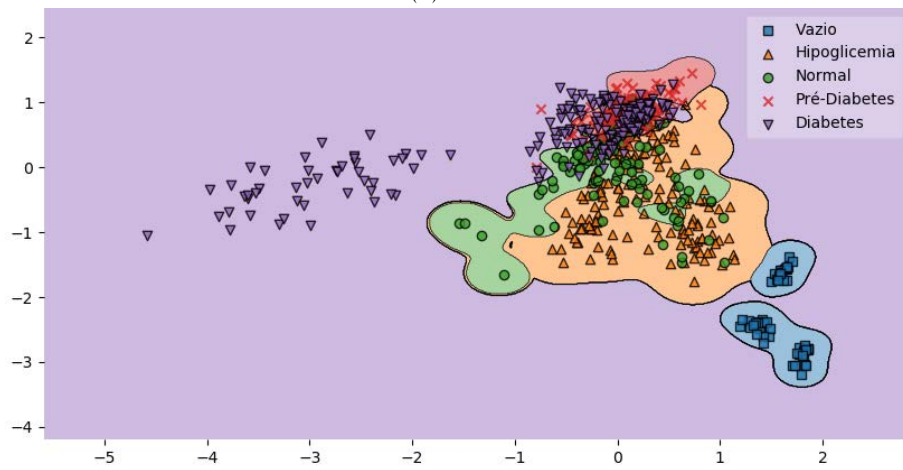
(c) *Random Forest*

Figura F.59: Regiões de decisão para os cinco níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados nos componentes S e V do espaço HSV.





(a) KNN



(b) SVM

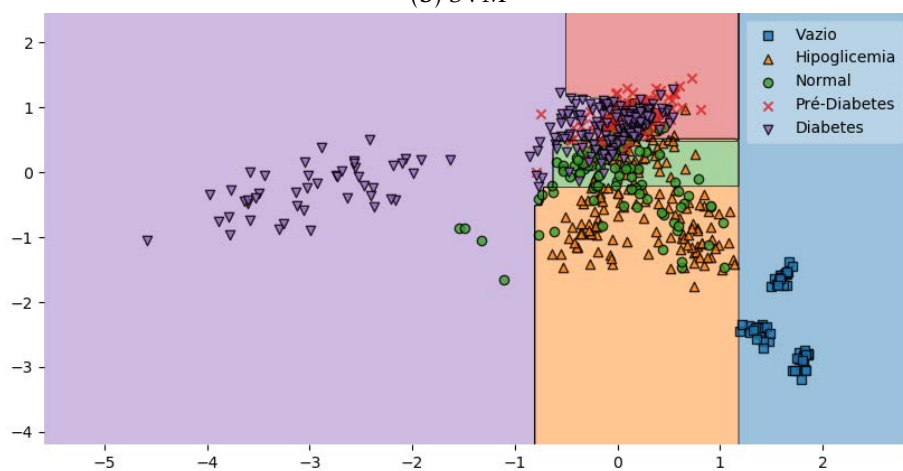
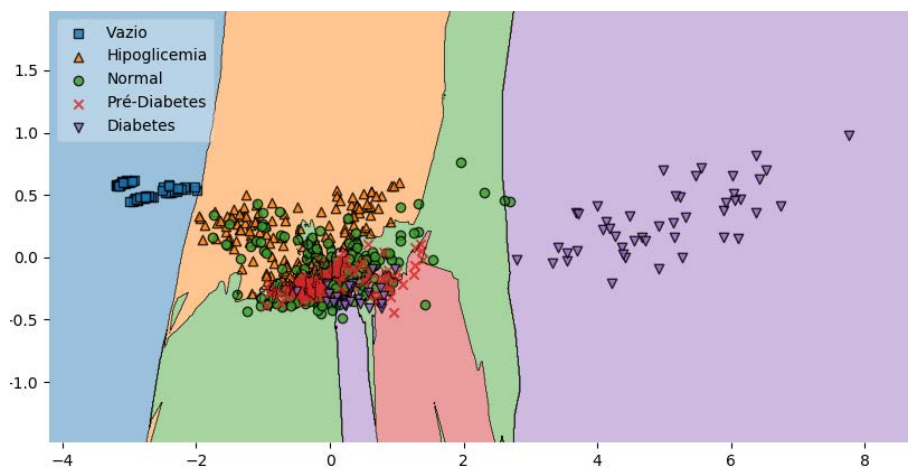
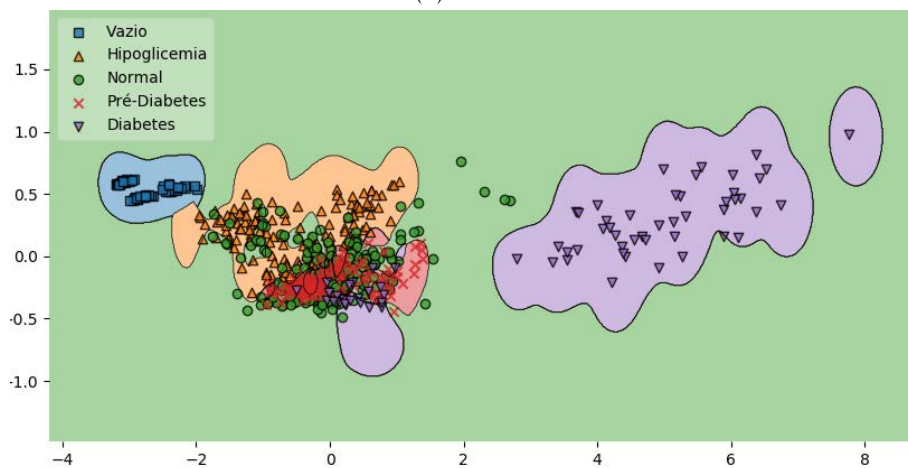
(c) *Random Forest*

Figura F.60: Regiões de decisão para os cinco níveis de referência da glucose (em jejum) dos três algoritmos estudados, baseados nos componentes L e S do espaço HLS.

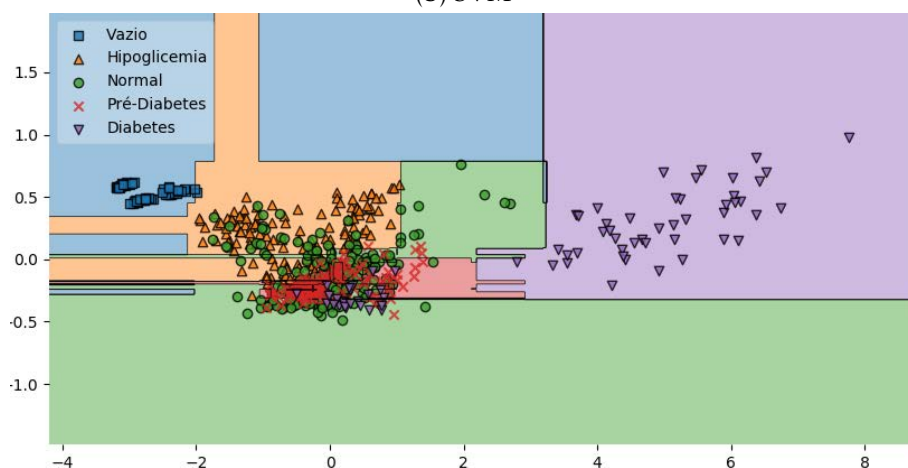
### F.3.2 Estado de duas horas após a refeição



(a) KNN



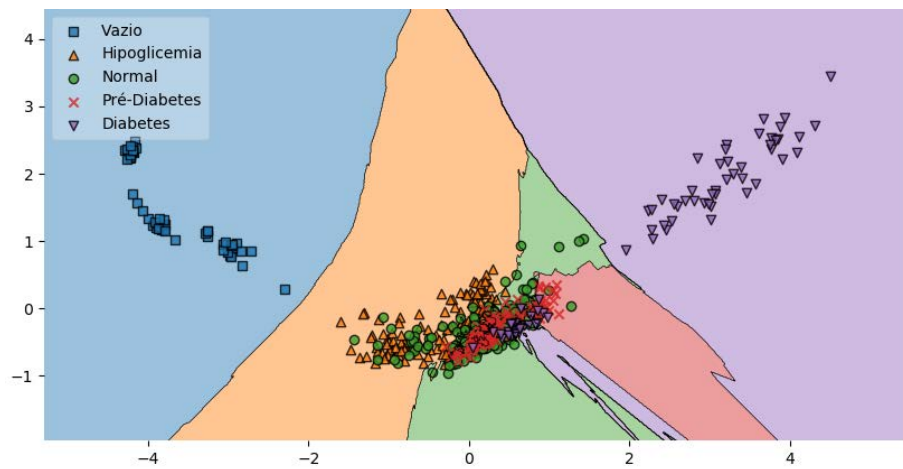
(b) SVM



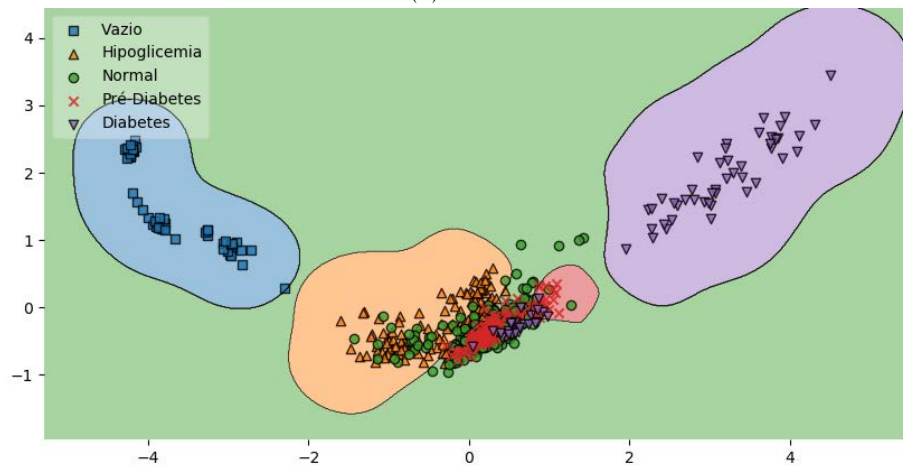
(c) Random Forest

Figura F.61: Regiões de decisão para os cinco níveis de referência da glicose (duas horas após a refeição) dos três algoritmos estudados, baseados na média dos componentes RGB.

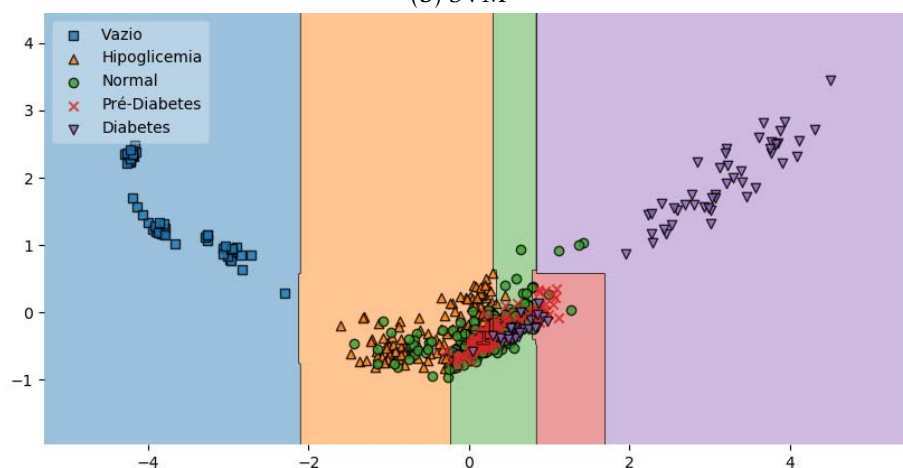




(a) KNN

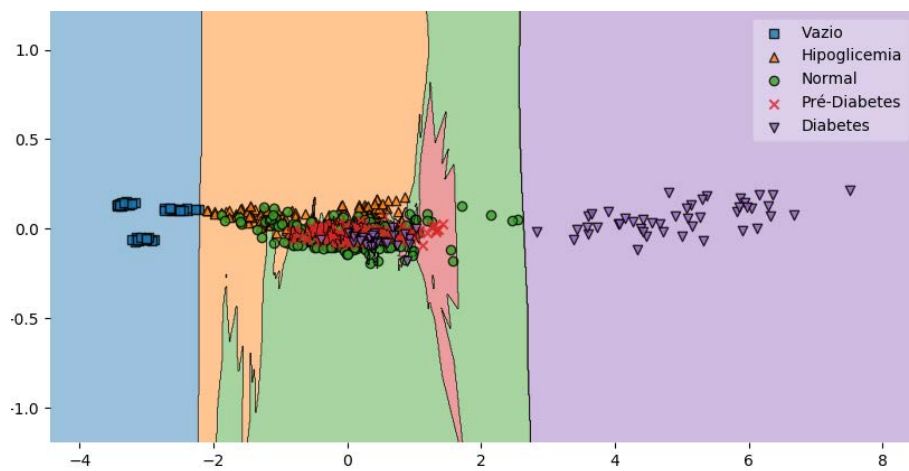


(b) SVM

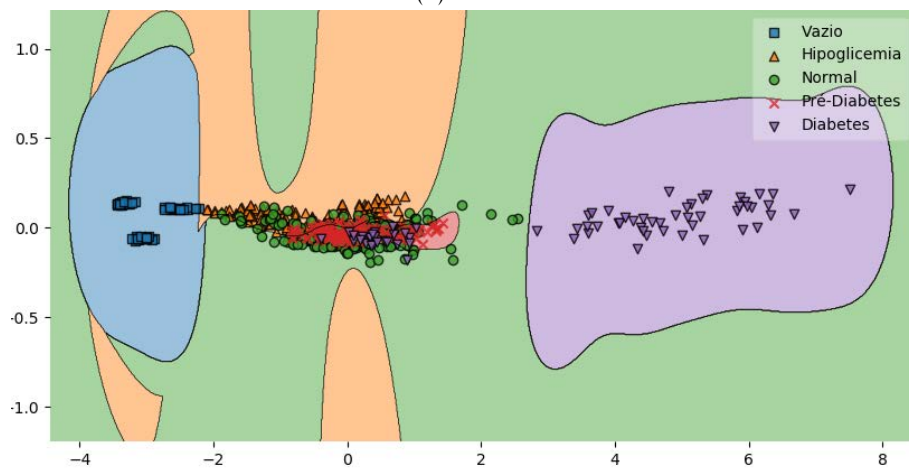


(c) Random Forest

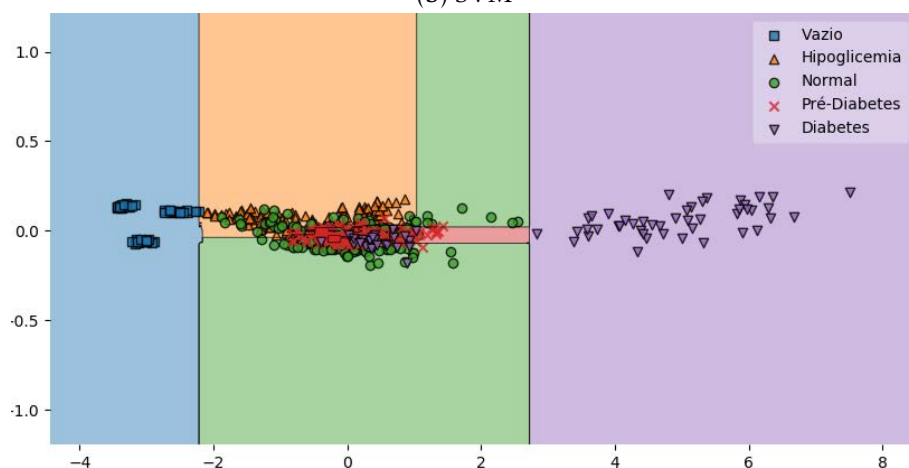
Figura F.62: Regiões de decisão para os cinco níveis de referência da glucose (duas horas após a refeição) dos três algoritmos estudados, baseados na média dos componentes HSV.



(a) KNN

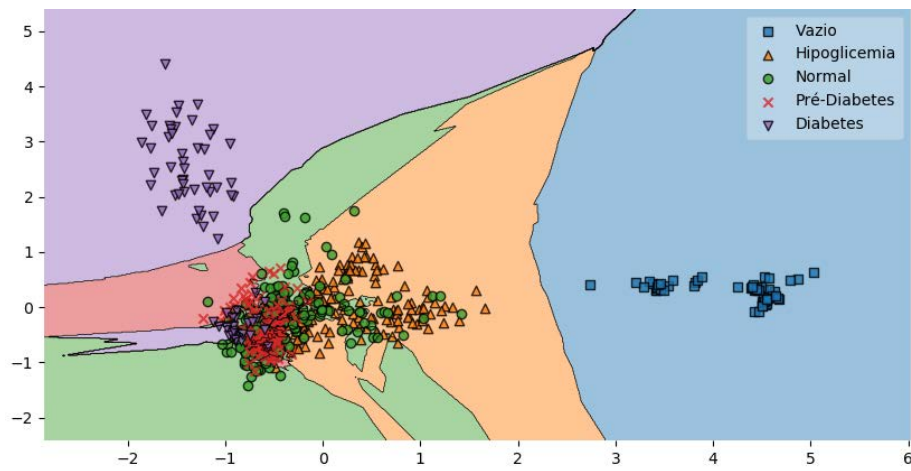


(b) SVM

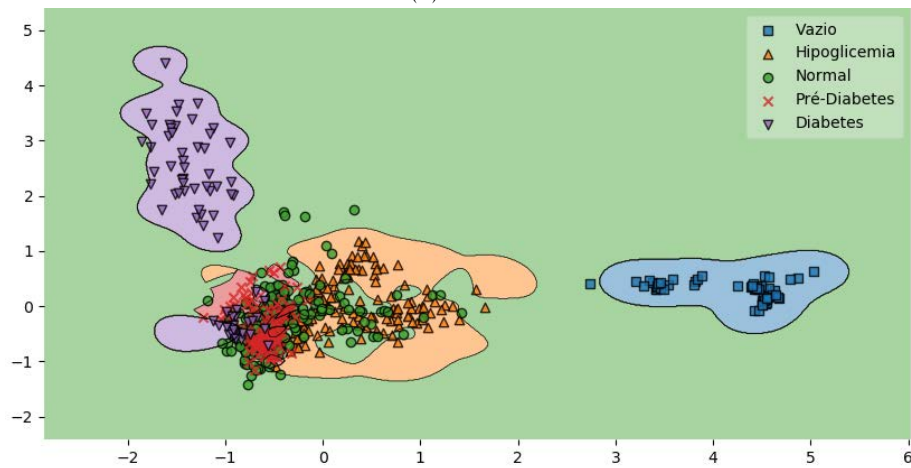


(c) *Random Forest*

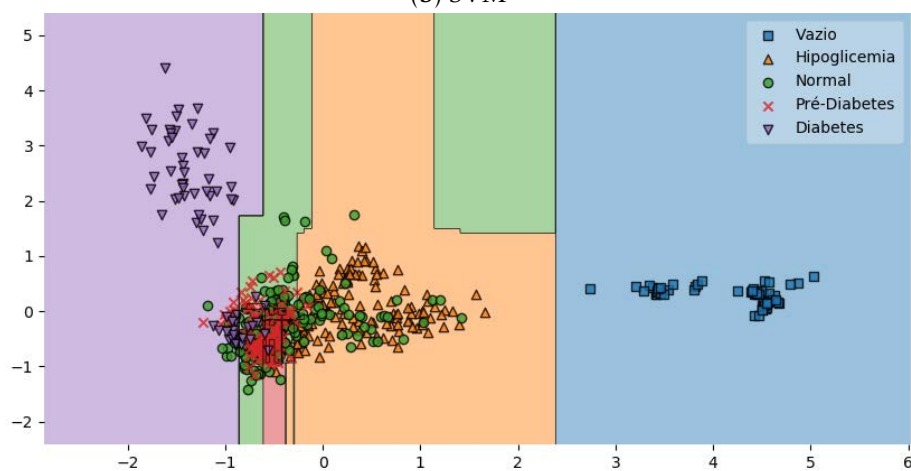
Figura F.63: Regiões de decisão para os cinco níveis de referência da glicose (duas horas após a refeição) dos três algoritmos estudados, baseados na média dos componentes XYZ.



(a) KNN

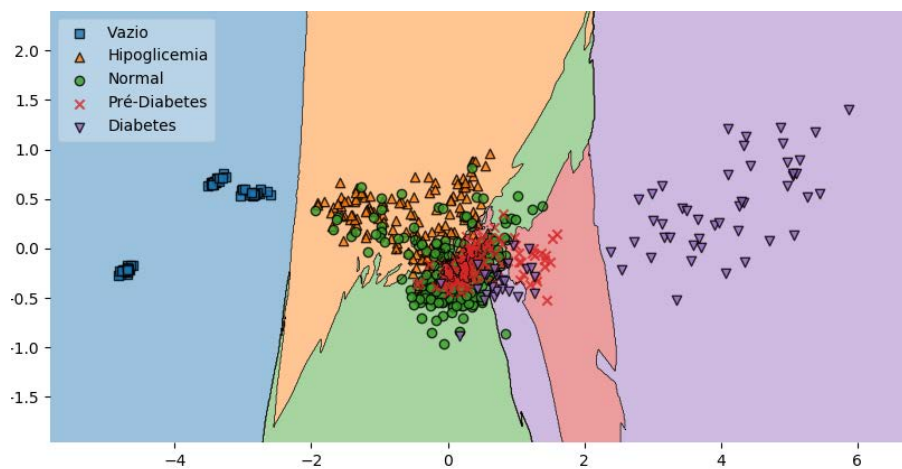


(b) SVM

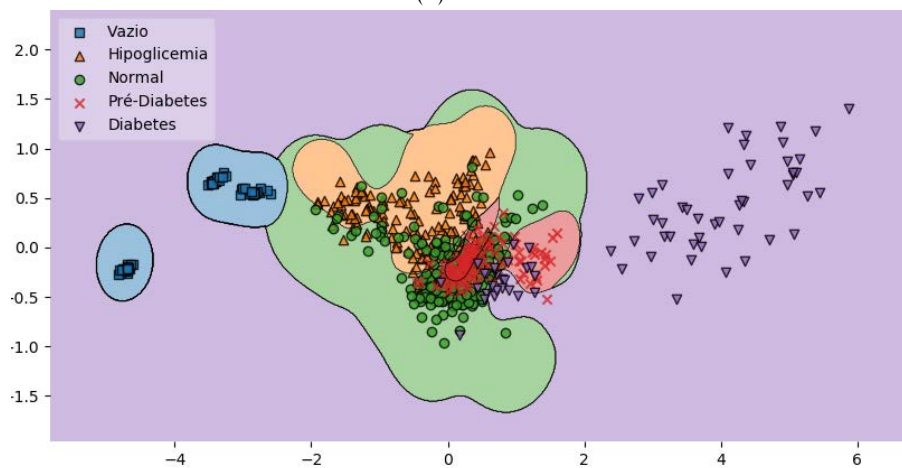


(c) Random Forest

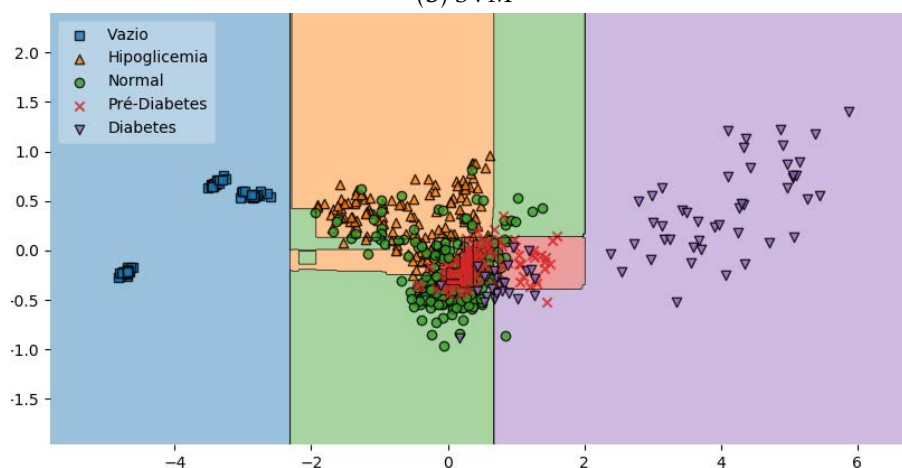
Figura F.64: Regiões de decisão para os cinco níveis de referência da glucose (duas horas após a refeição) dos três algoritmos estudados, baseados na média dos componentes HLS.



(a) KNN

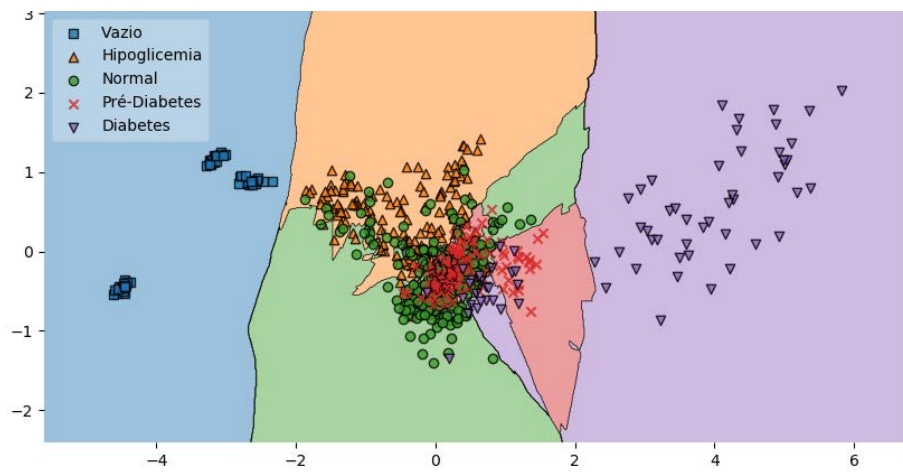


(b) SVM

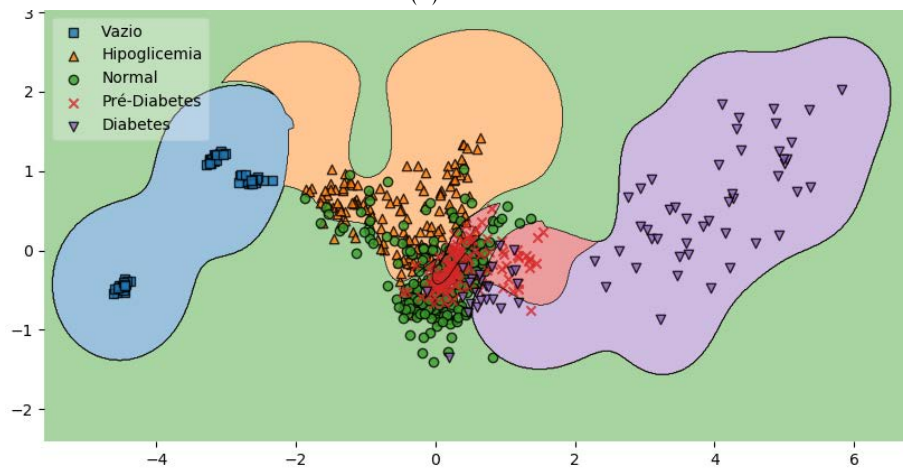


(c) Random Forest

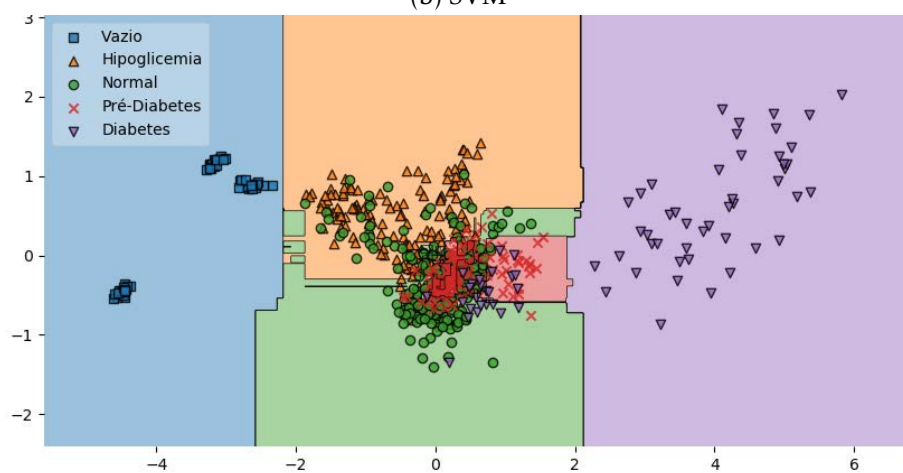
Figura F.65: Regiões de decisão para os cinco níveis de referência da glicose (duas horas após a refeição) dos três algoritmos estudados, baseados na média dos componentes LAB.



(a) KNN

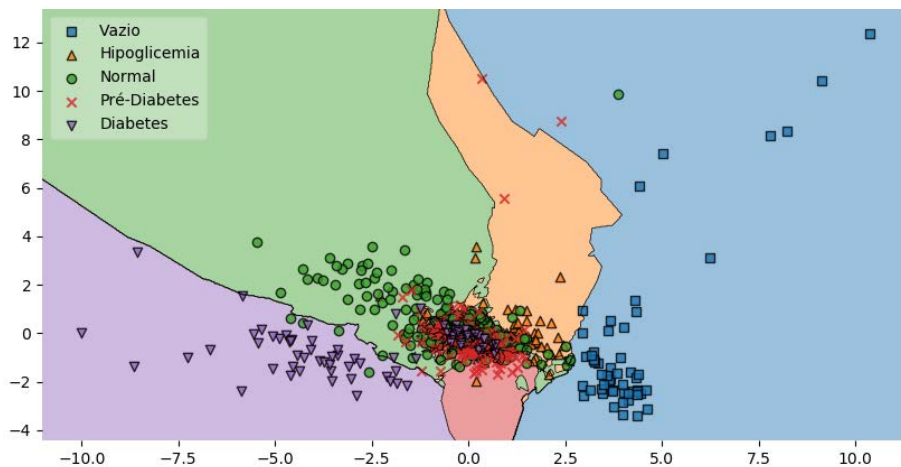


(b) SVM

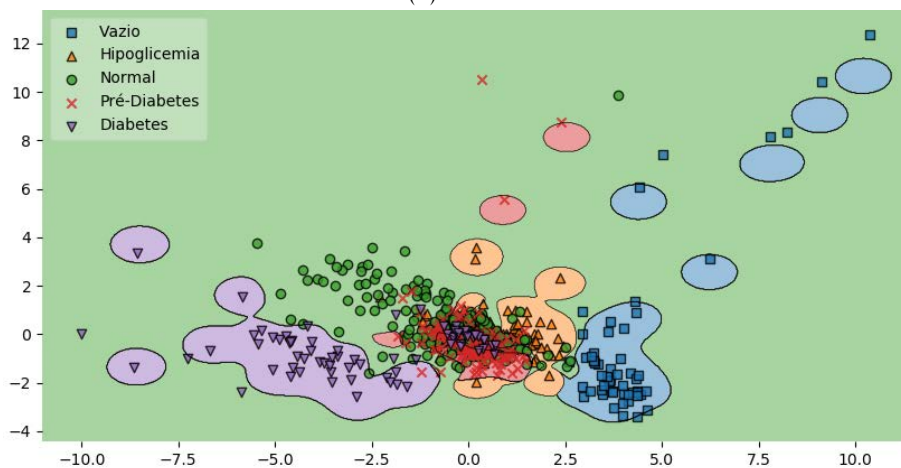


(c) Random Forest

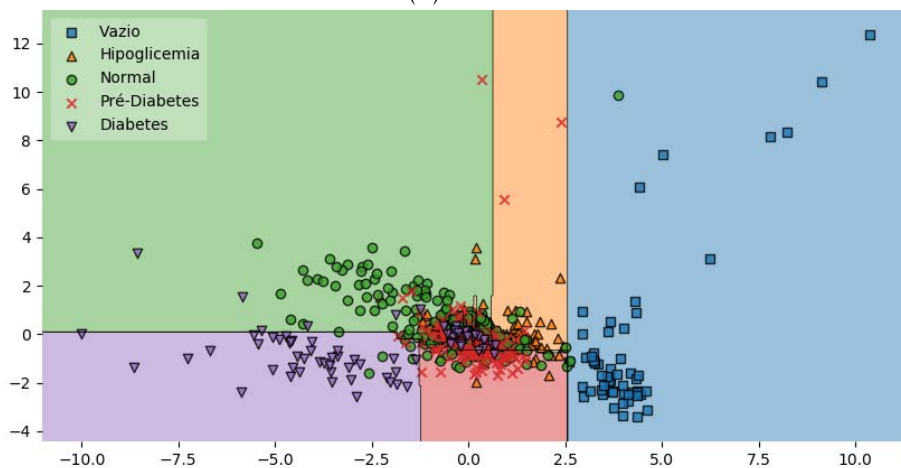
Figura F.66: Regiões de decisão para os cinco níveis de referência da glucose (duas horas após a refeição) dos três algoritmos estudados, baseados na média dos componentes YUV.



(a) KNN



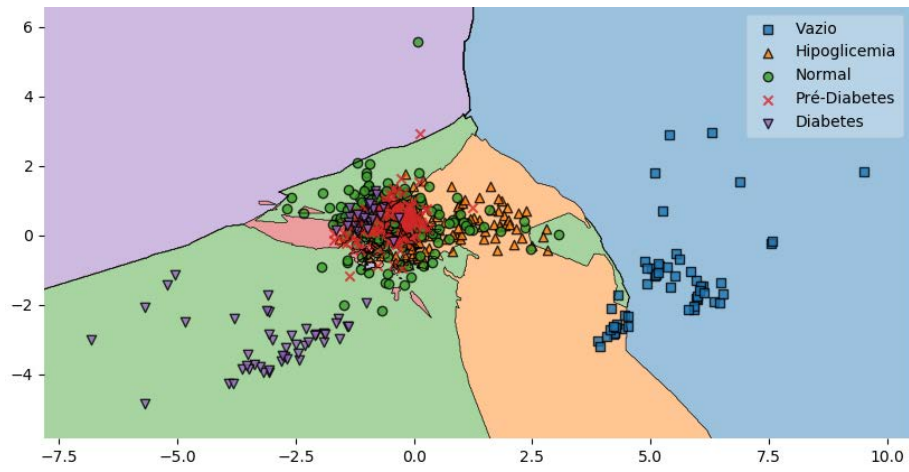
(b) SVM



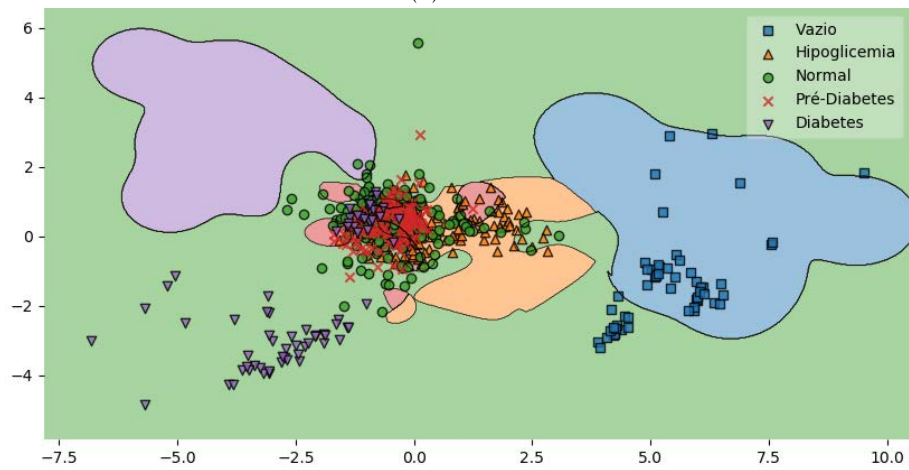
(c) Random Forest

Figura F.67: Regiões de decisão para os cinco níveis de referência da glicose (duas horas após a refeição) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes RGB.

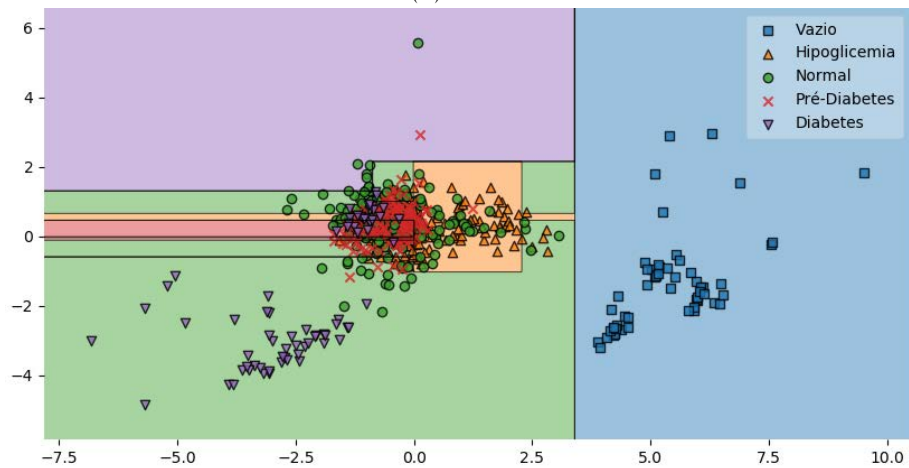




(a) KNN

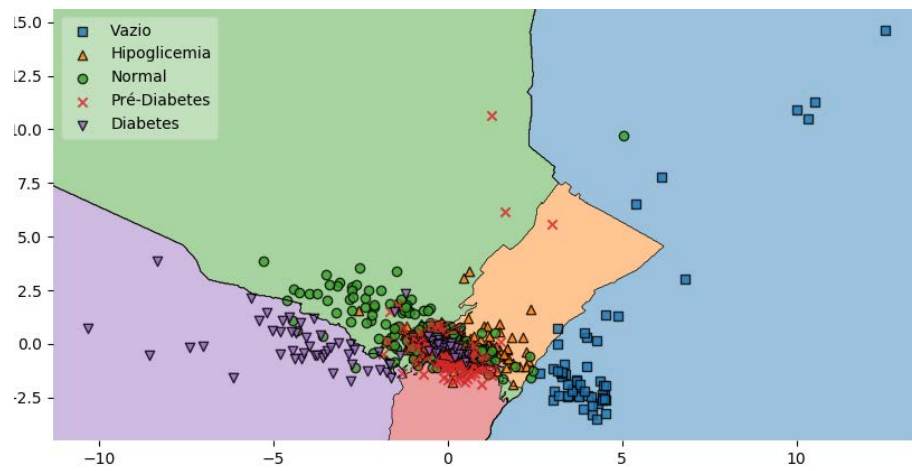


(b) SVM

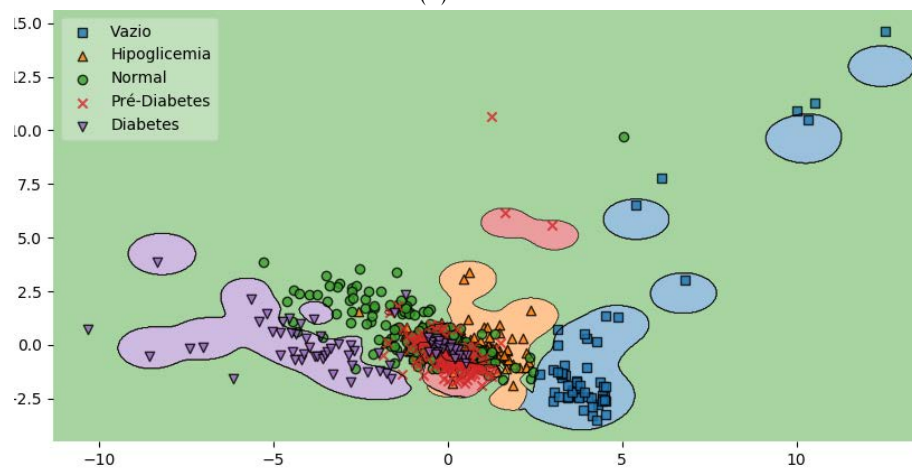


(c) Random Forest

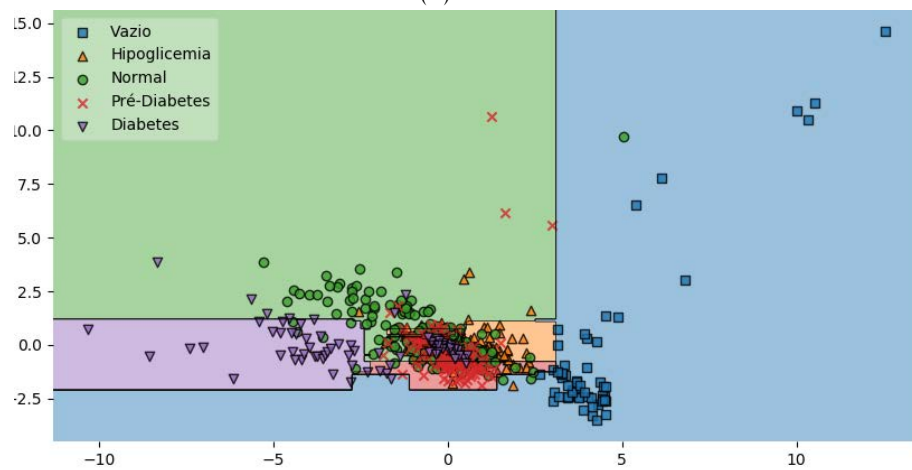
Figura F.68: Regiões de decisão para os cinco níveis de referência da glucose (duas horas após a refeição) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes HSV.



(a) KNN



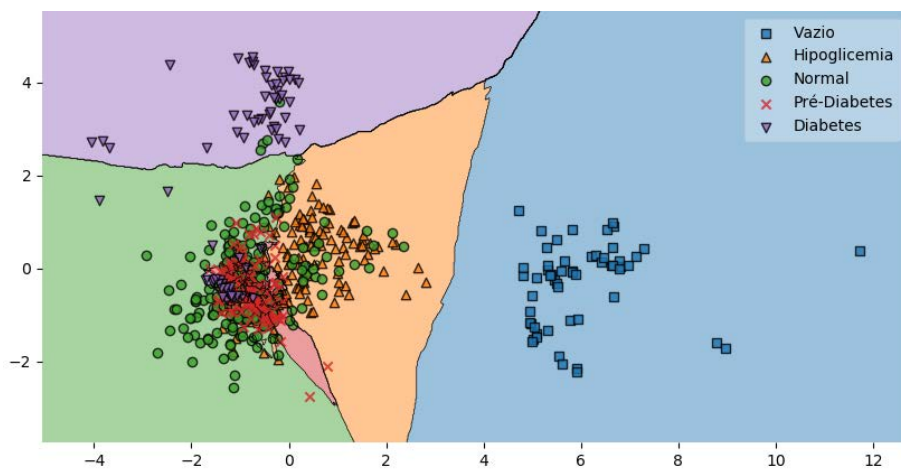
(b) SVM



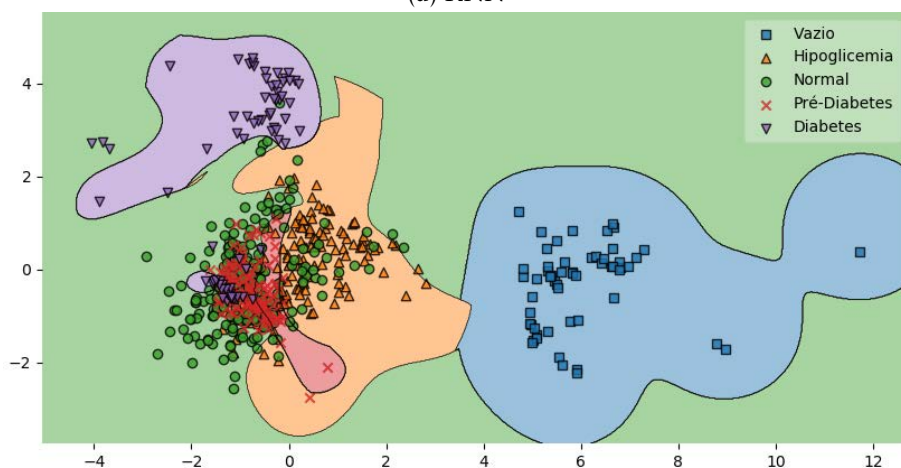
(c) Random Forest

Figura F.69: Regiões de decisão para os cinco níveis de referência da glicose (duas horas após a refeição) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes XYZ.





(a) KNN



(b) SVM

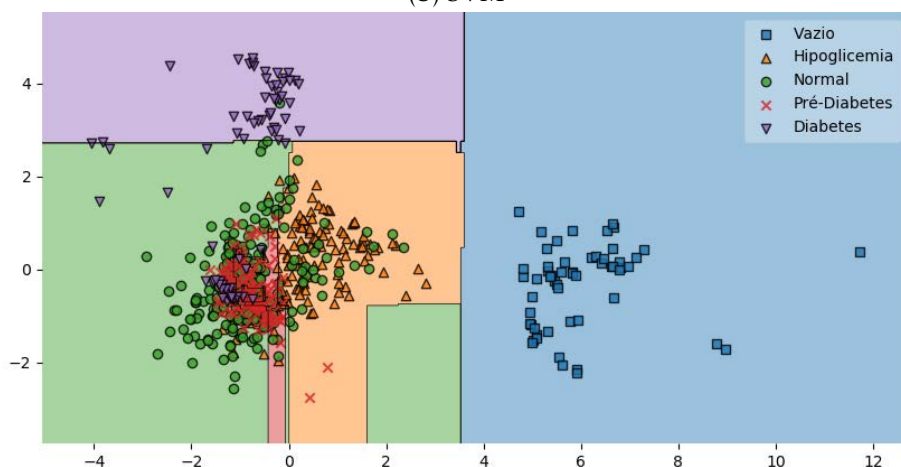
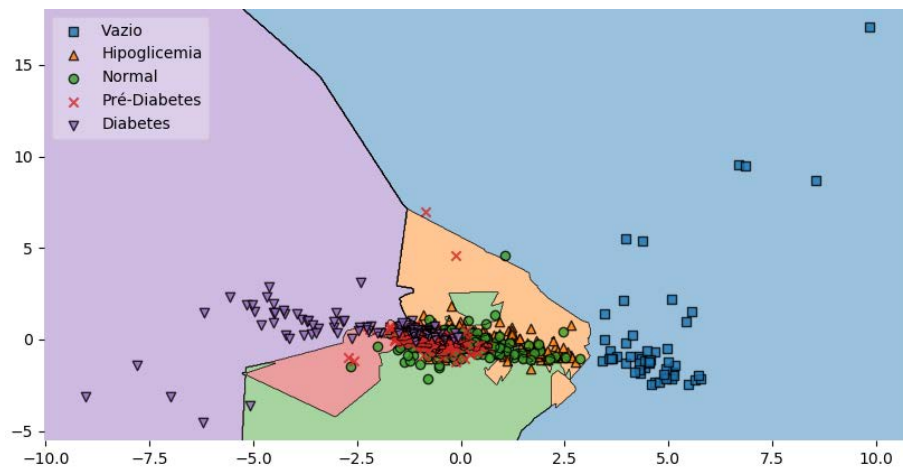
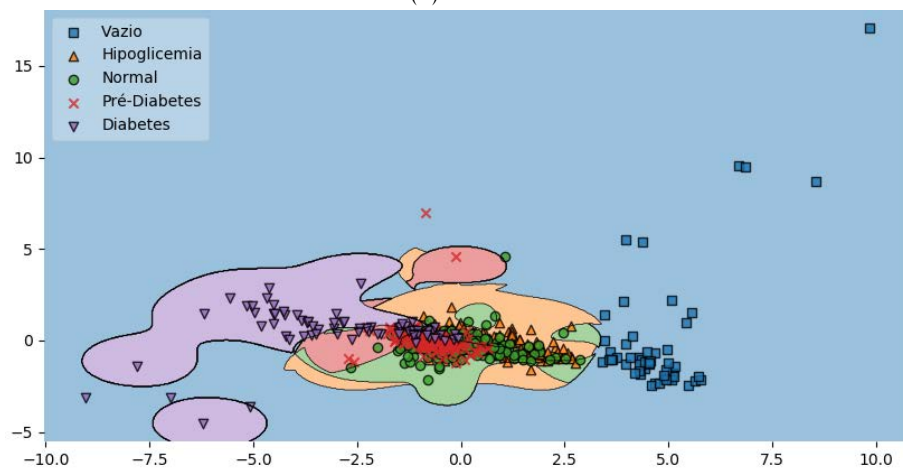
(c) *Random Forest*

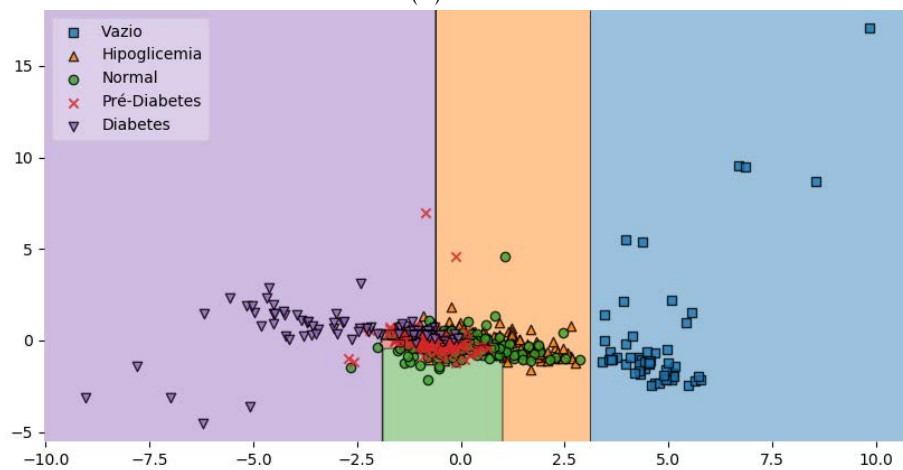
Figura F.70: Regiões de decisão para os cinco níveis de referência da glucose (duas horas após a refeição) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes HLS.



(a) KNN

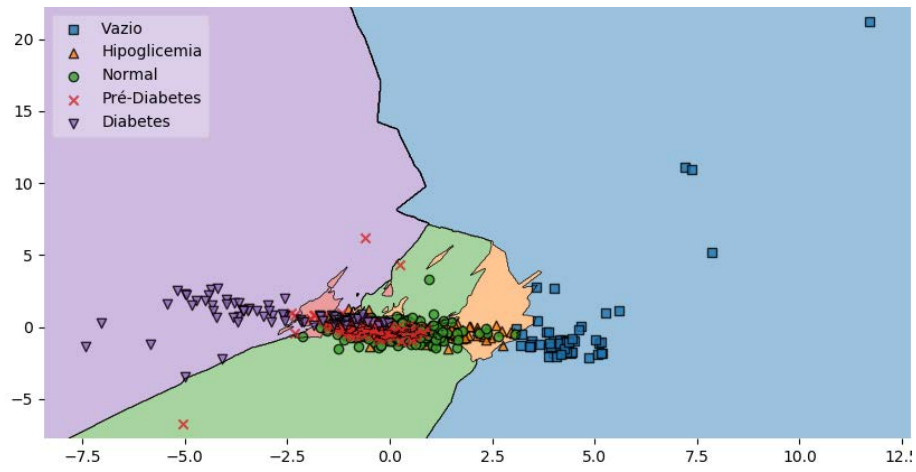


(b) SVM

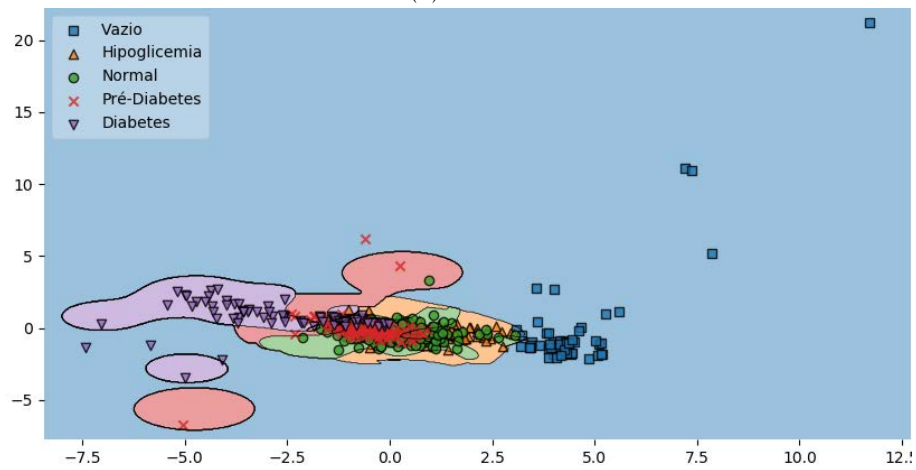


(c) *Random Forest*

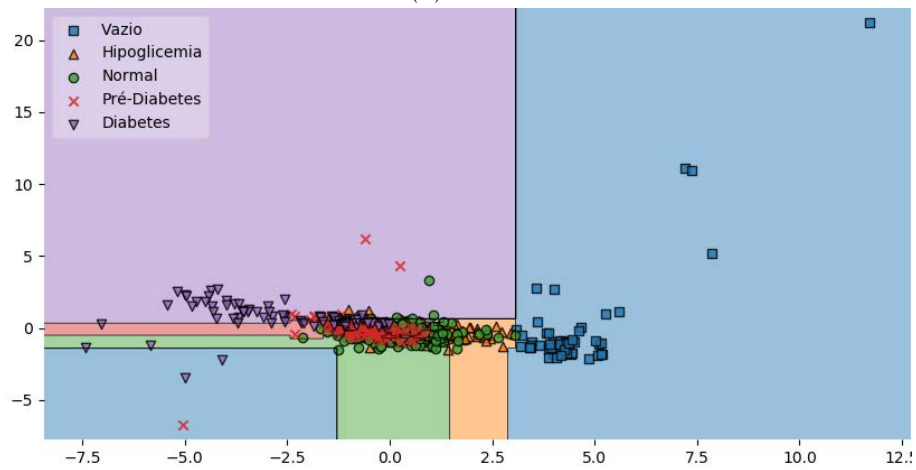
Figura F.71: Regiões de decisão para os cinco níveis de referência da glicose (duas horas após a refeição) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes LAB.



(a) KNN

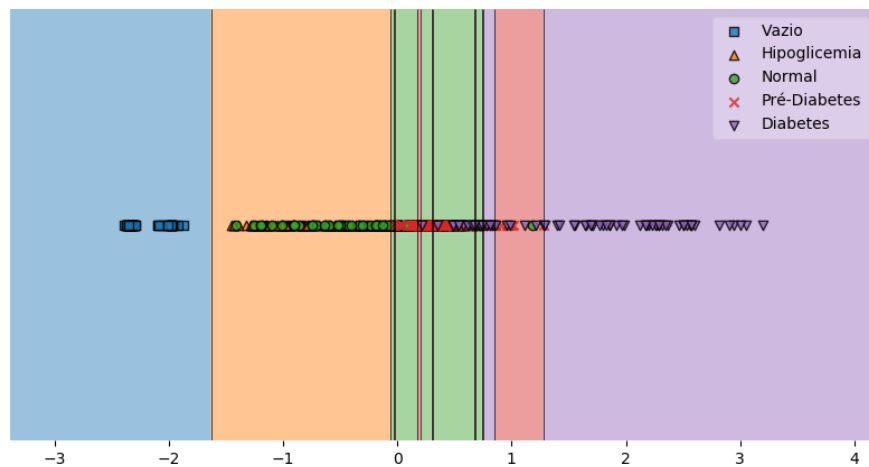


(b) SVM

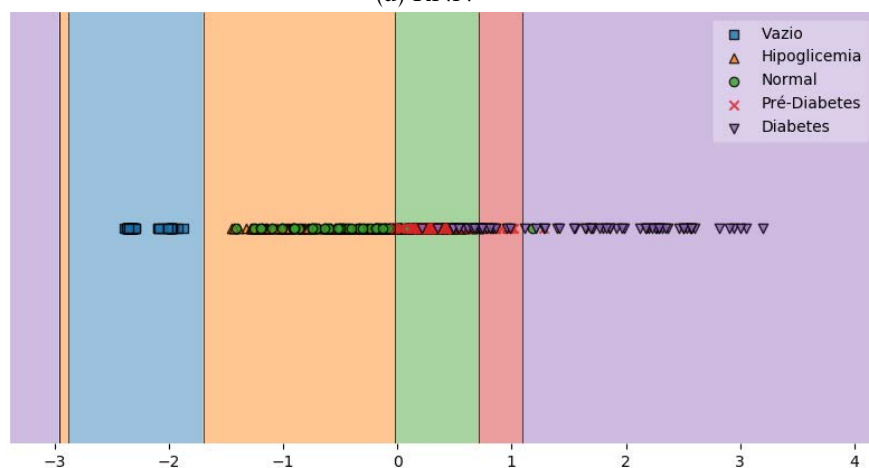


(c) Random Forest

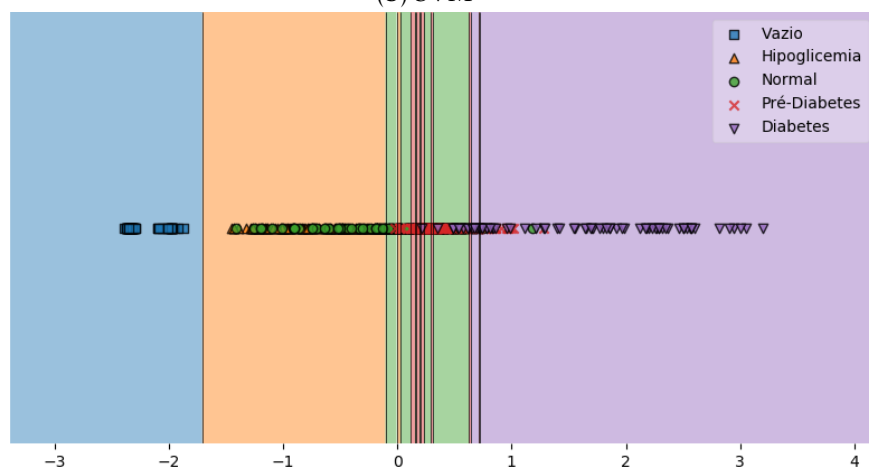
Figura F.72: Regiões de decisão para os cinco níveis de referência da glucose (duas horas após a refeição) dos três algoritmos estudados, baseados na média, desvio padrão e assimetria dos componentes YUV.



(a) KNN

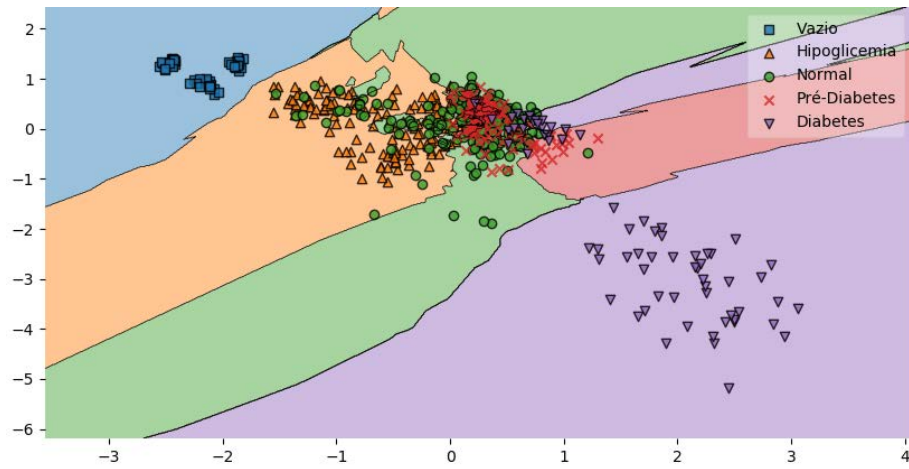


(b) SVM

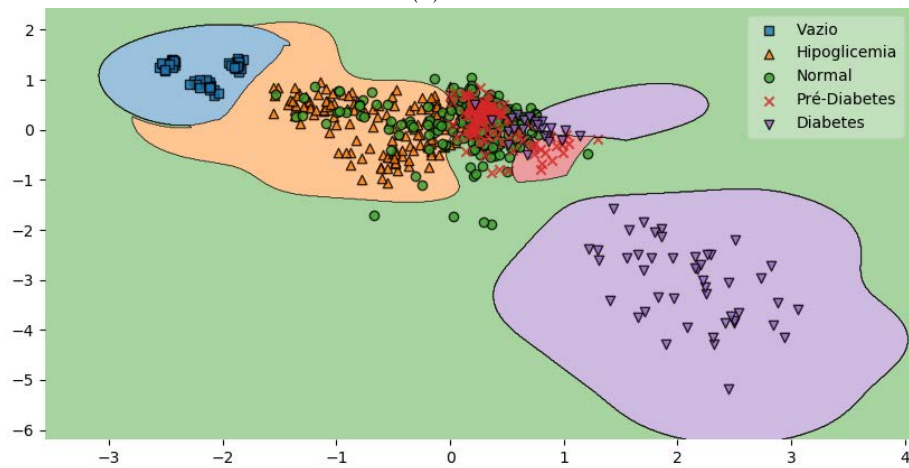


(c) Random Forest

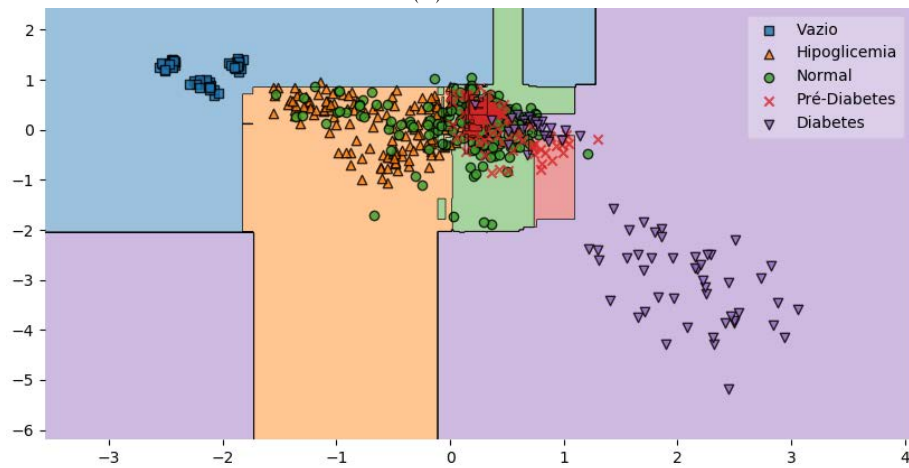
Figura F.73: Regiões de decisão para os cinco níveis de referência da glicose (duas horas após a refeição) dos três algoritmos estudados, baseados no rácio entre os componentes vermelho e verde do espaço RGB.



(a) KNN

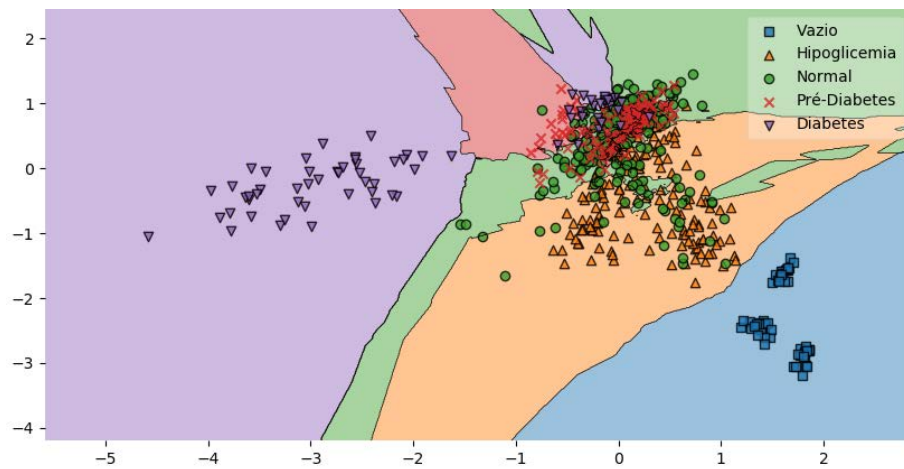


(b) SVM

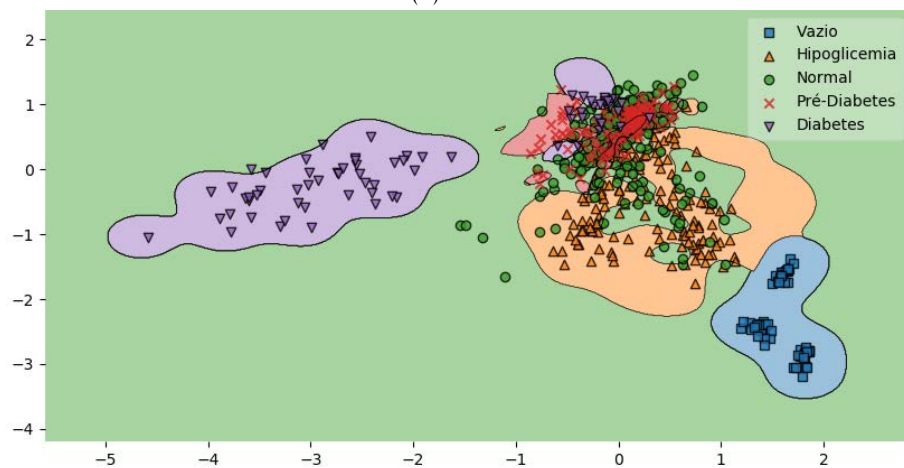


(c) Random Forest

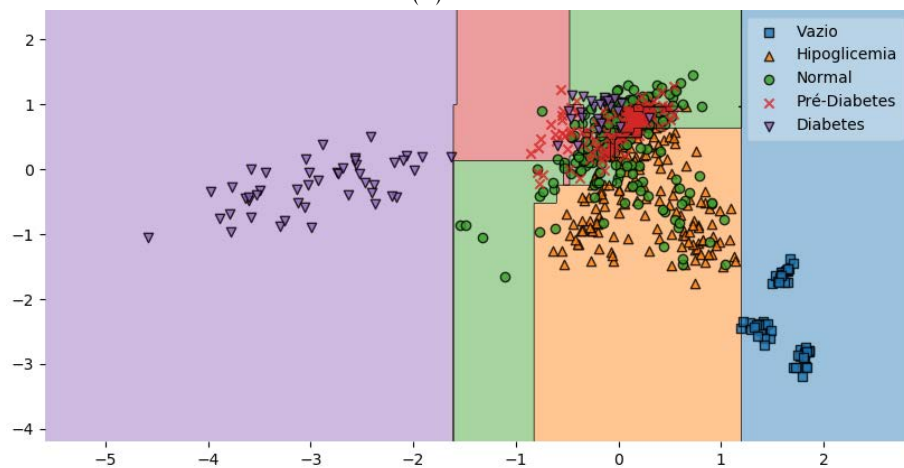
Figura F.74: Regiões de decisão para os cinco níveis de referência da glucose (duas horas após a refeição) dos três algoritmos estudados, baseados nos componentes S e V do espaço HSV.



(a) KNN



(b) SVM

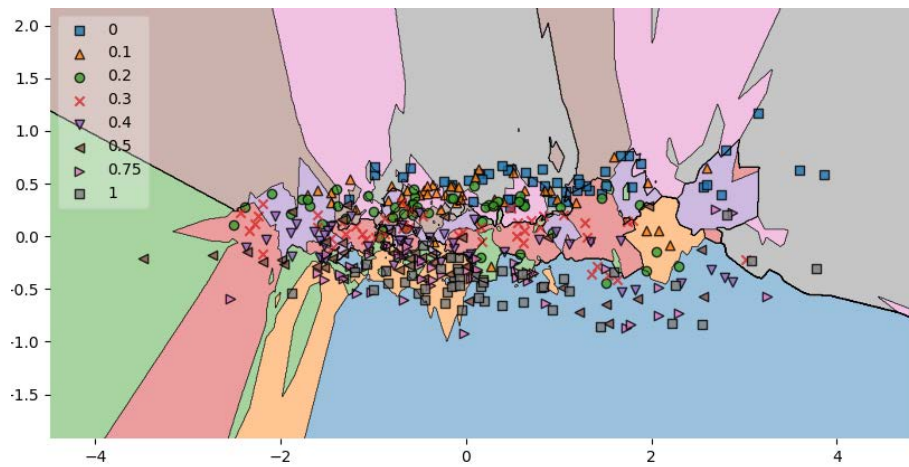


(c) Random Forest

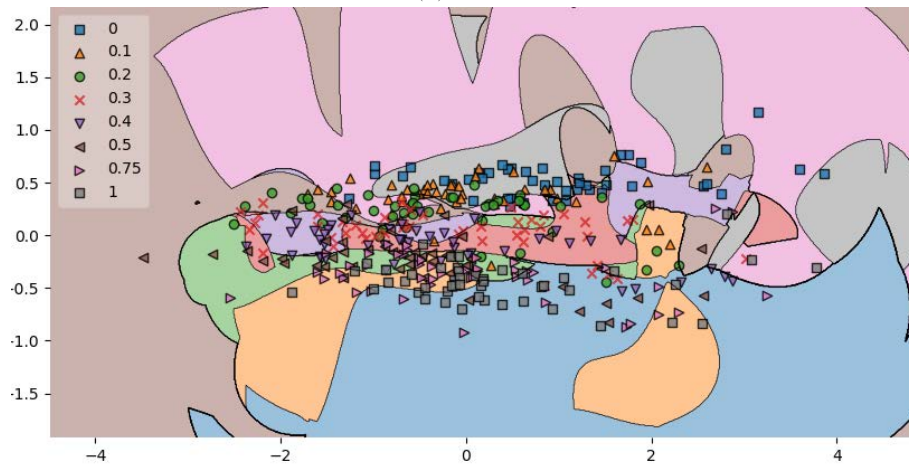
Figura F.75: Regiões de decisão para os cinco níveis de referência da glicose (duas horas após a refeição) dos três algoritmos estudados, baseados nos componentes L e S do espaço HLS.



## F.4 Concentrações de Ácido Úrico



(a) KNN



(b) SVM

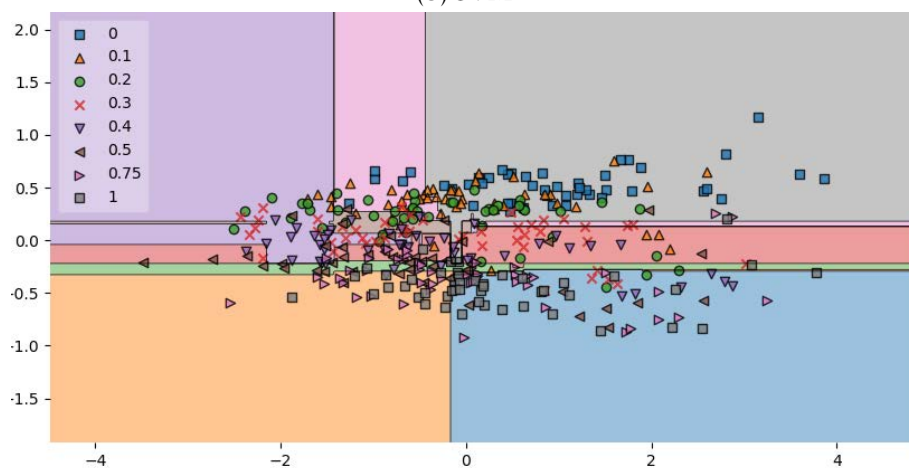
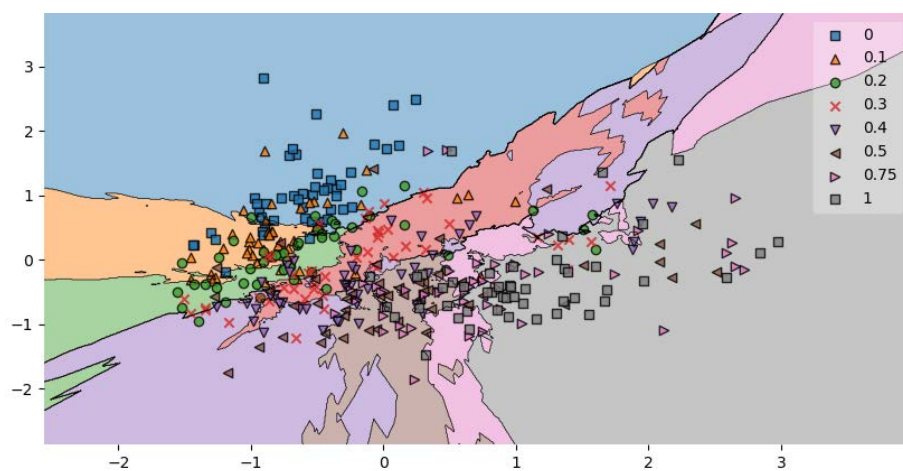
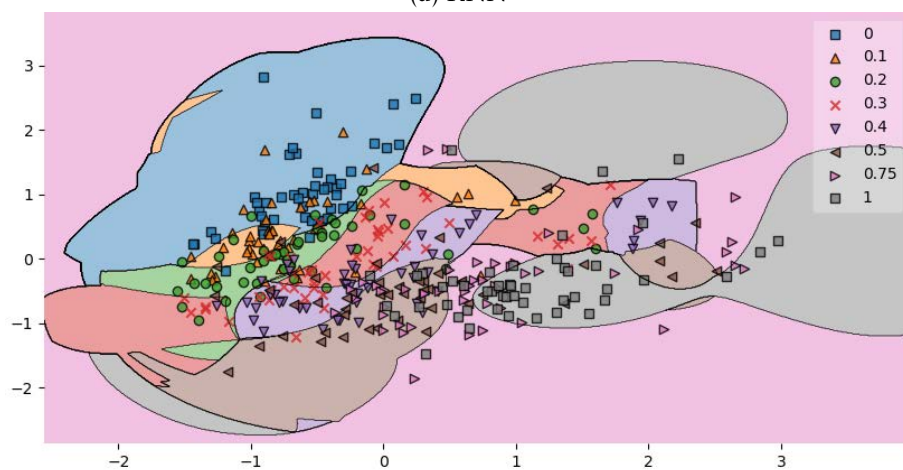
(c) *Random Forest*

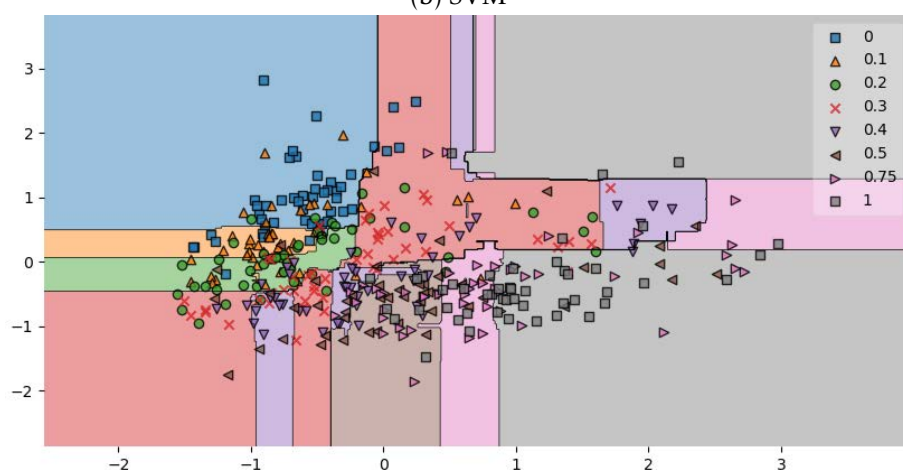
Figura F.76: Regiões de decisão para as concentrações do ácido úrico (mM) dos três algoritmos estudados, baseados na média dos componentes RGB.



(a) KNN



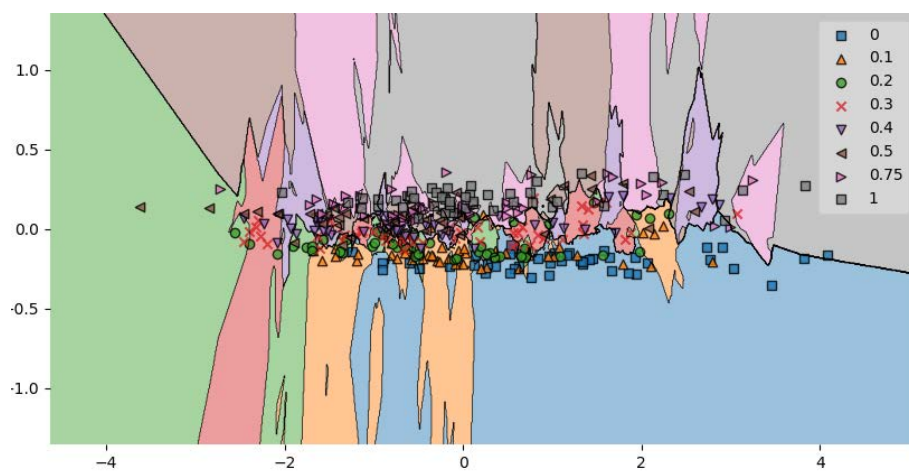
(b) SVM



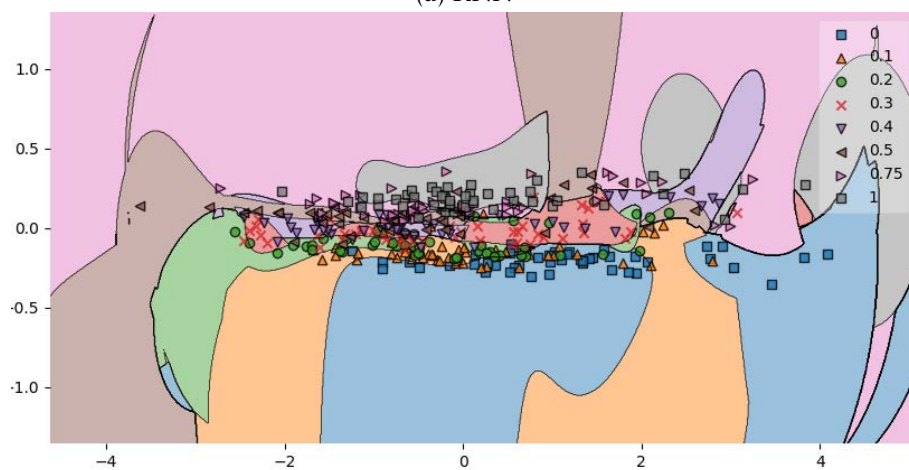
(c) *Random Forest*

Figura F.77: Regiões de decisão para as concentrações do ácido úrico (mM) dos três algoritmos estudados, baseados na média dos componentes HSV.

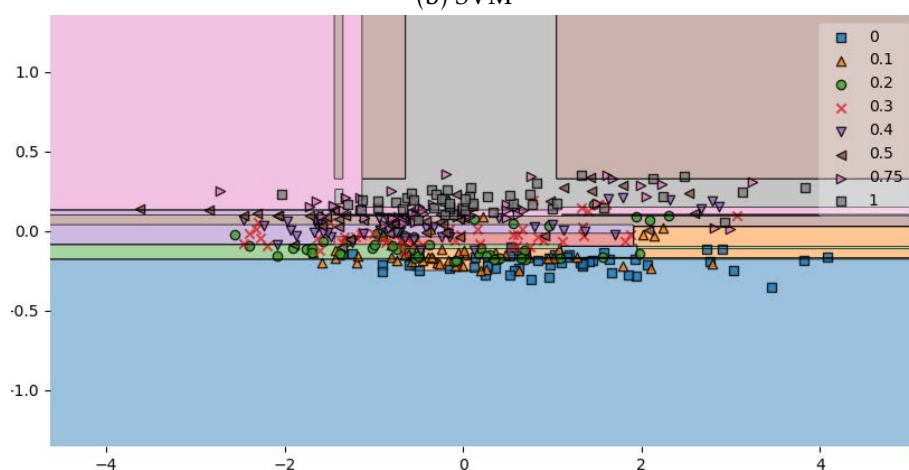




(a) KNN

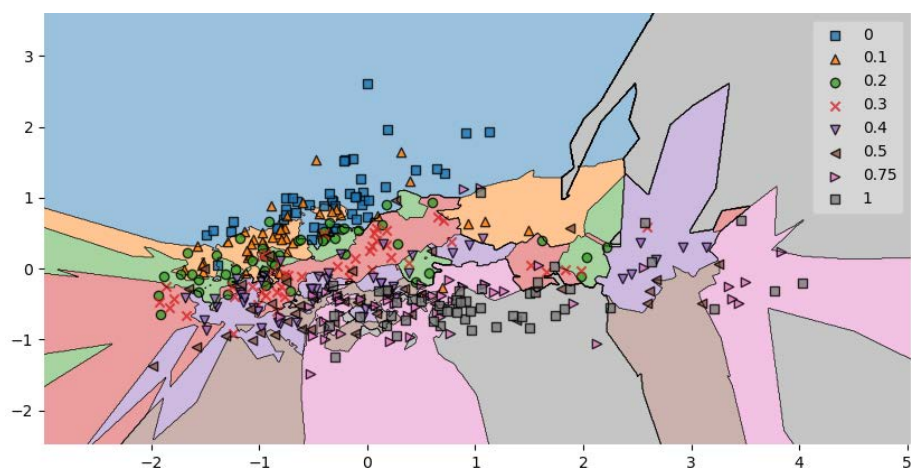


(b) SVM

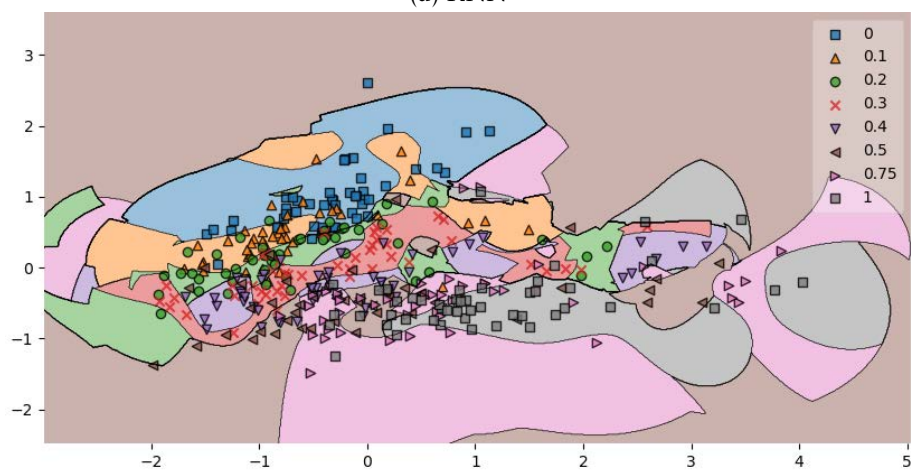


(c) *Random Forest*

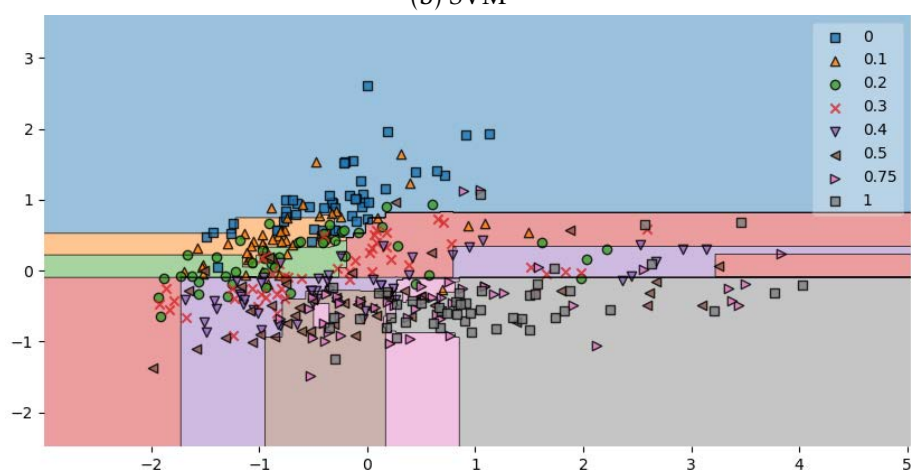
Figura F.78: Regiões de decisão para as concentrações do ácido úrico (mM) dos três algoritmos estudados, baseados na média dos componentes XYZ.



(a) KNN

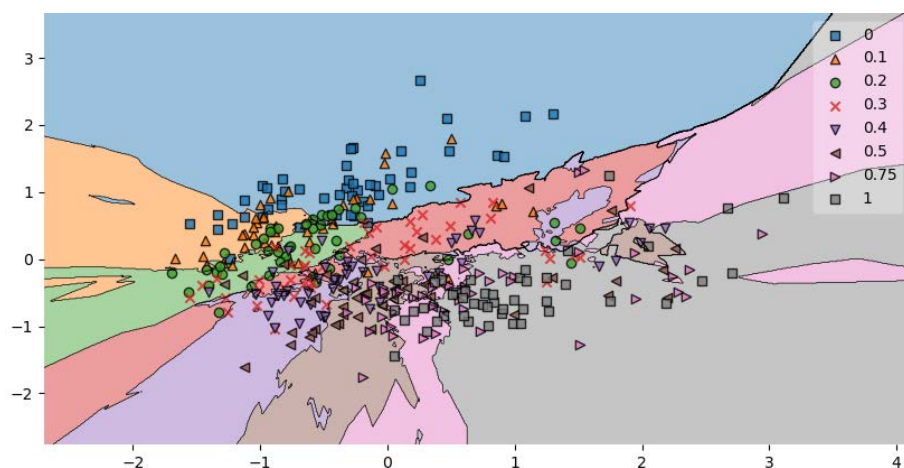


(b) SVM

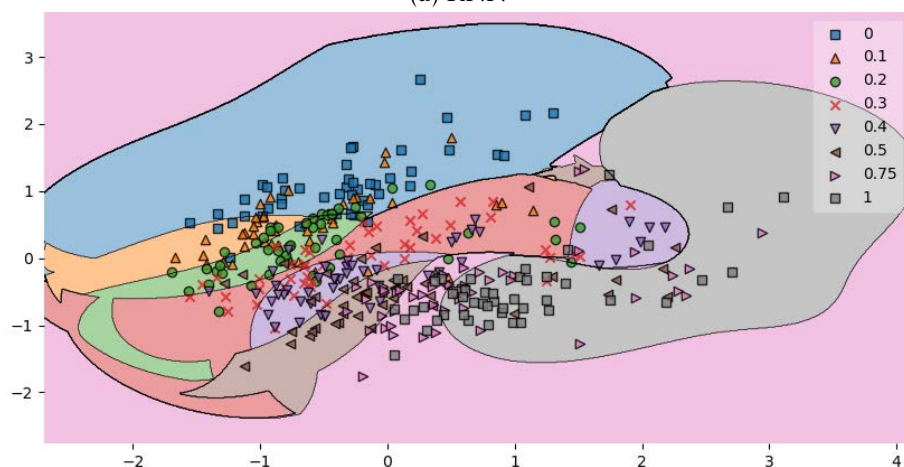


(c) *Random Forest*

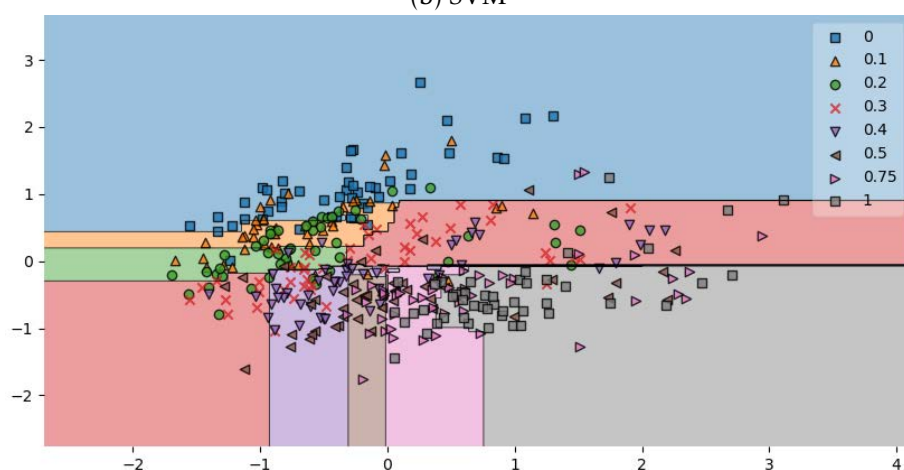
Figura F.79: Regiões de decisão para as concentrações do ácido úrico (mM) dos três algoritmos estudados, baseados na média dos componentes HLS.



(a) KNN

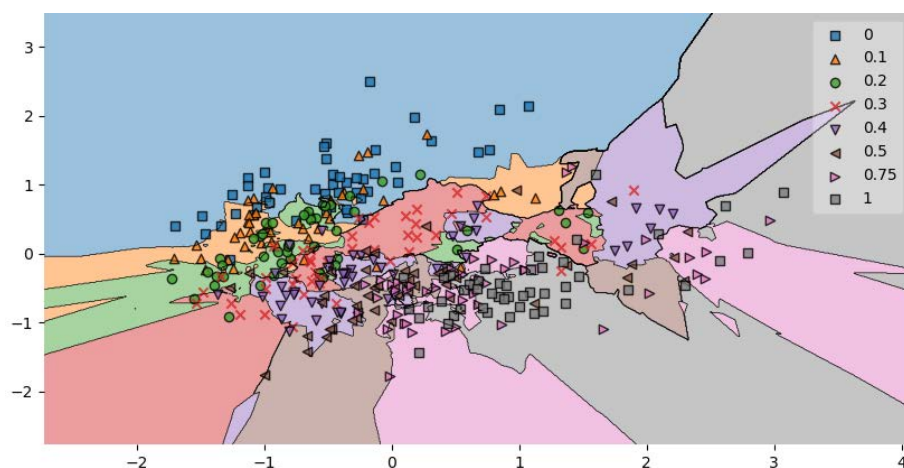


(b) SVM

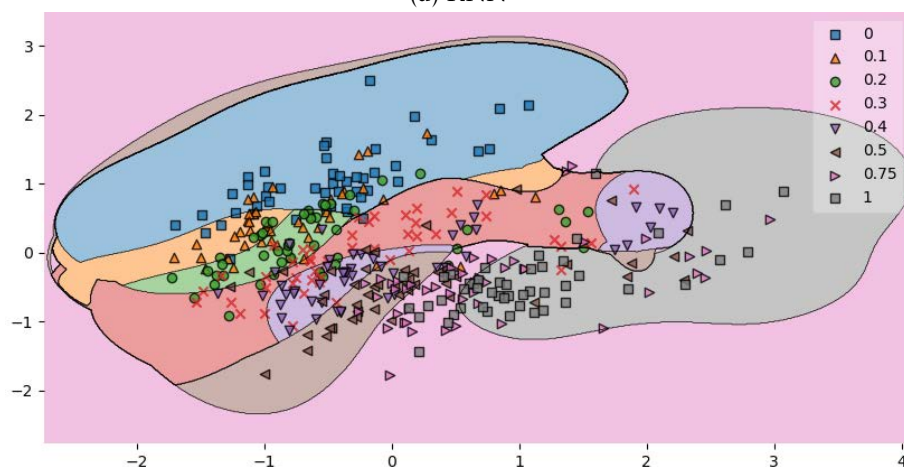


(c) *Random Forest*

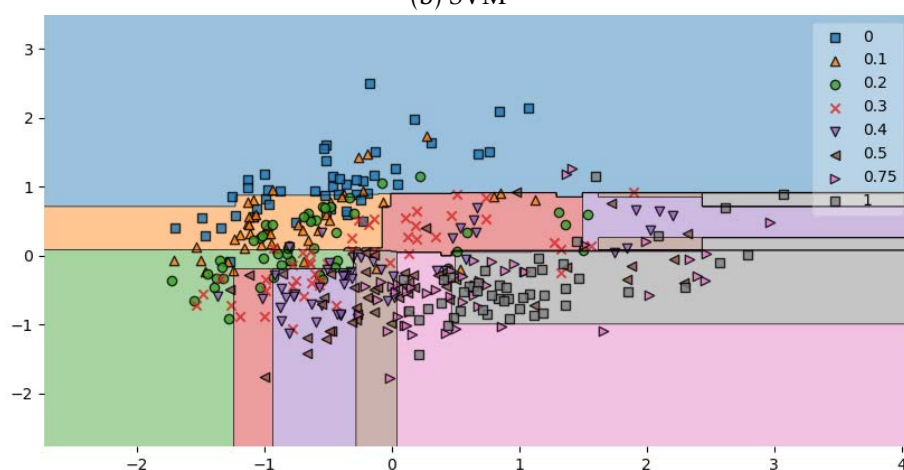
Figura F.80: Regiões de decisão para as concentrações do ácido úrico (mM) dos três algoritmos estudados, baseados na média dos componentes LAB.



(a) KNN

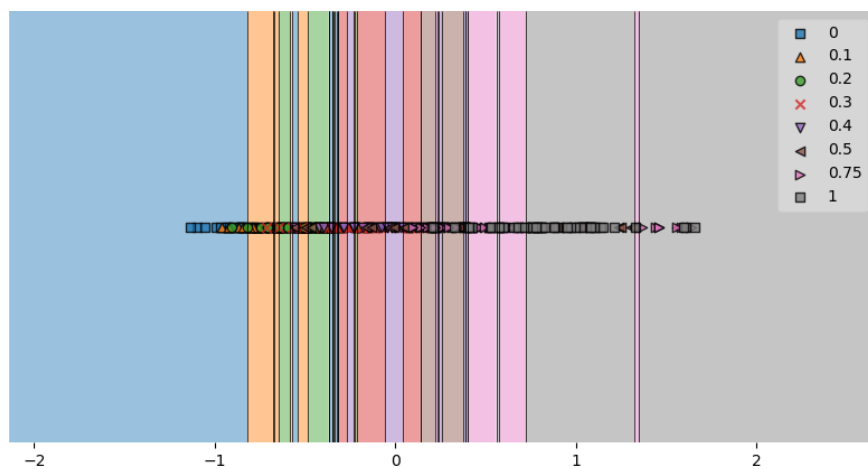


(b) SVM

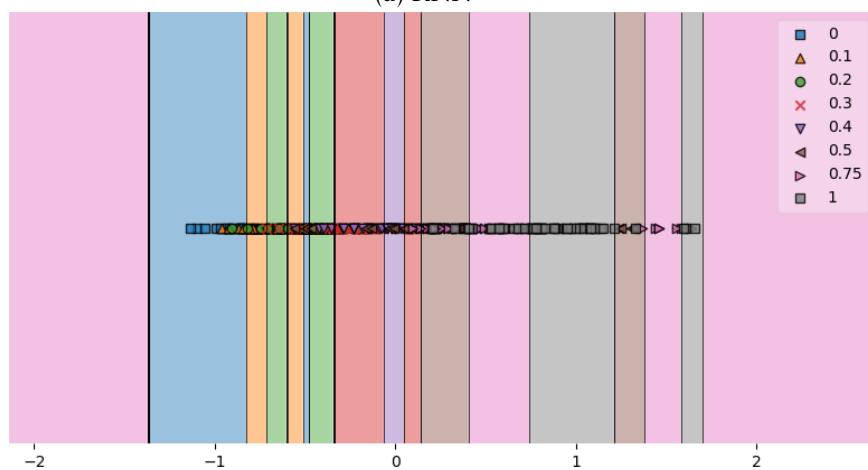


(c) *Random Forest*

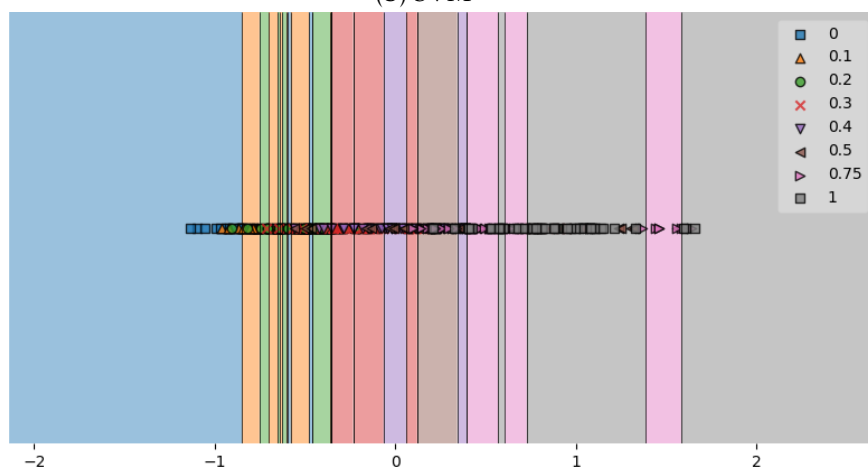
Figura F.81: Regiões de decisão para as concentrações do ácido úrico (mM) dos três algoritmos estudados, baseados na média dos componentes YUV.



(a) KNN

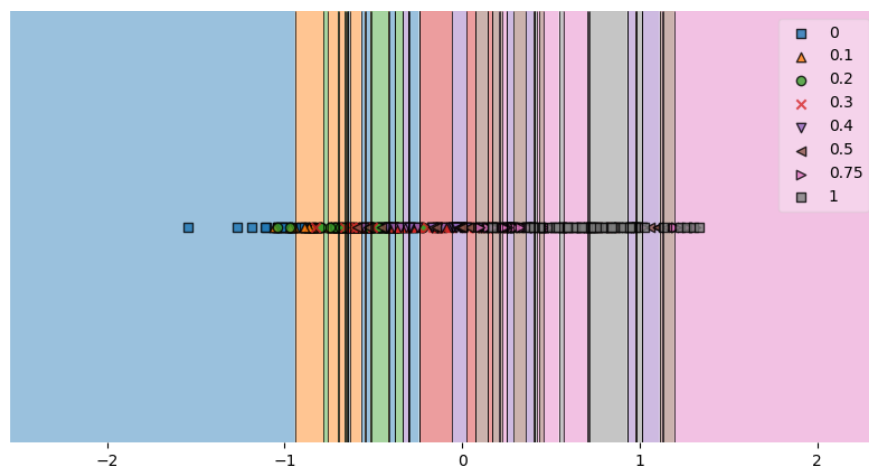


(b) SVM

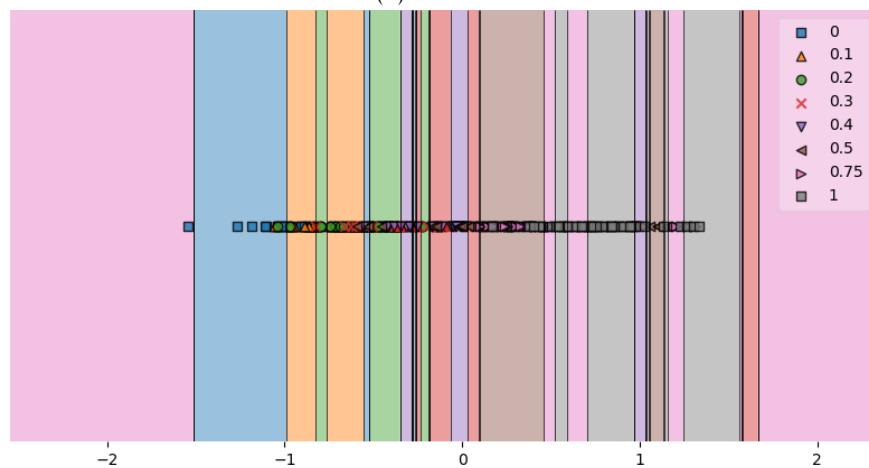


(c) Random Forest

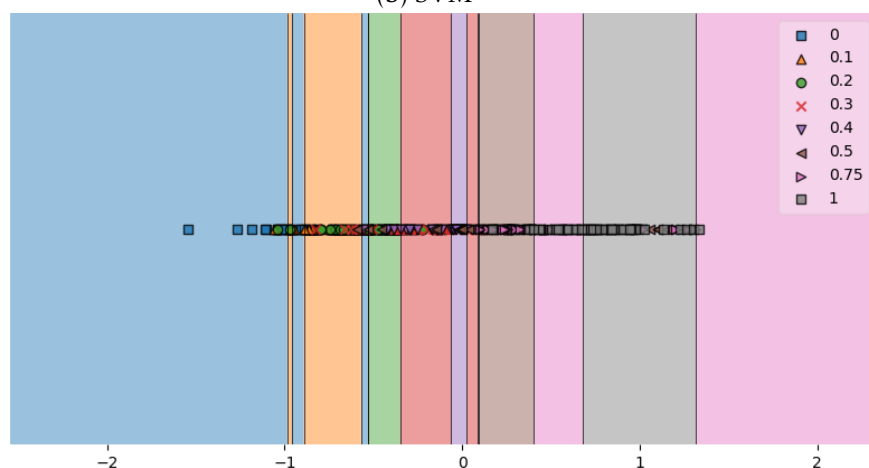
Figura F.82: Regiões de decisão para as concentrações do ácido úrico (mM) dos três algoritmos estudados, baseados no rácio entre os componentes vermelho e azul do espaço RGB.



(a) KNN



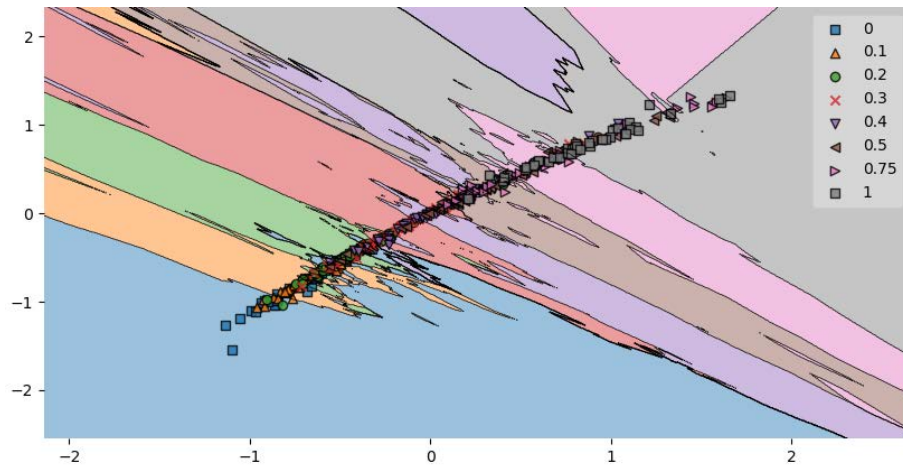
(b) SVM



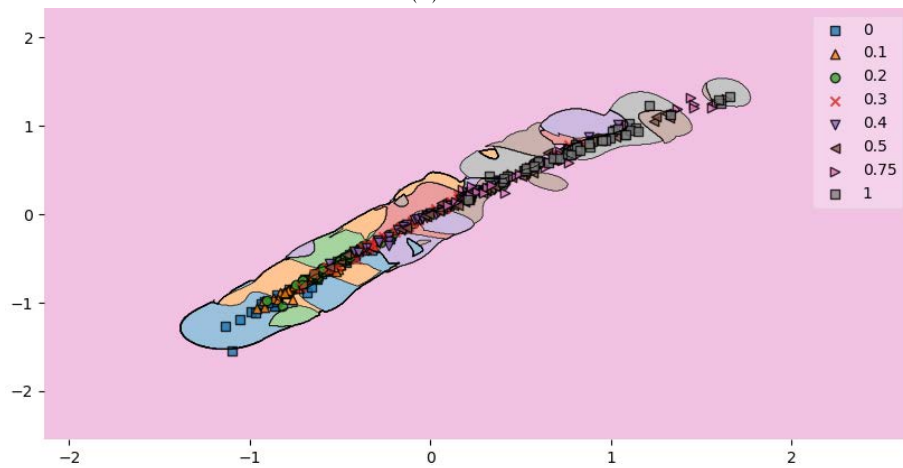
(c) *Random Forest*

Figura F.83: Regiões de decisão para as concentrações do ácido úrico (mM) dos três algoritmos estudados, baseados no componente H do espaço HSV.

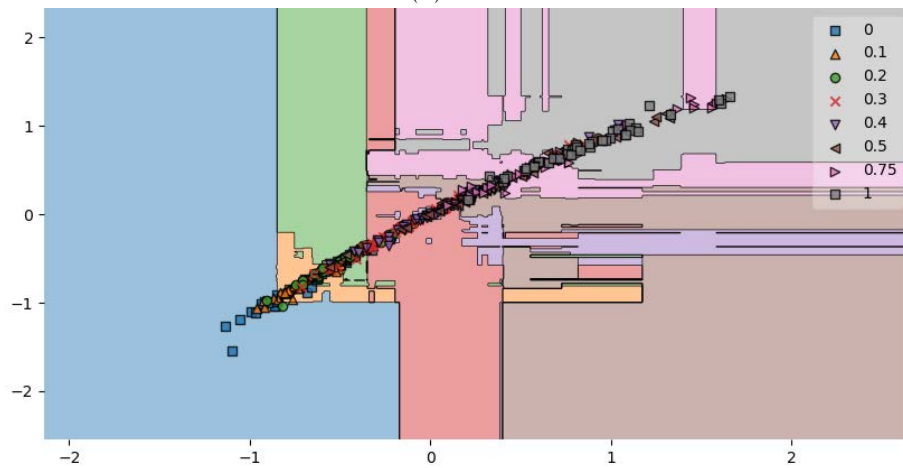




(a) KNN

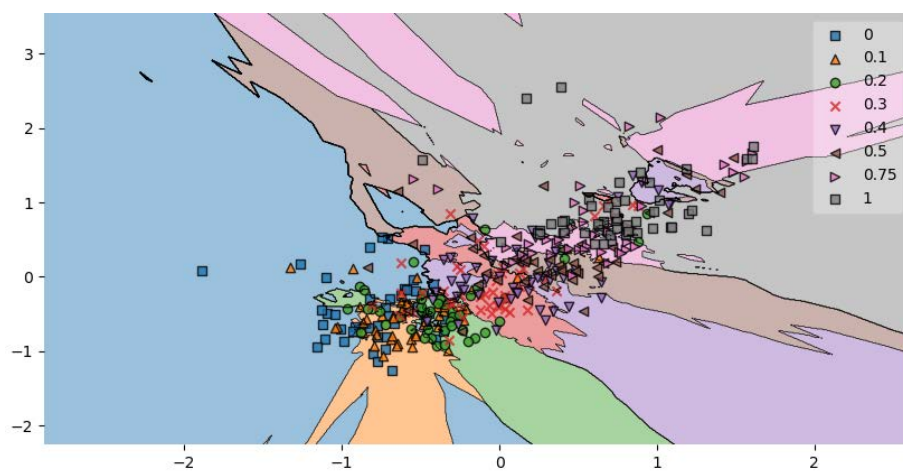


(b) SVM

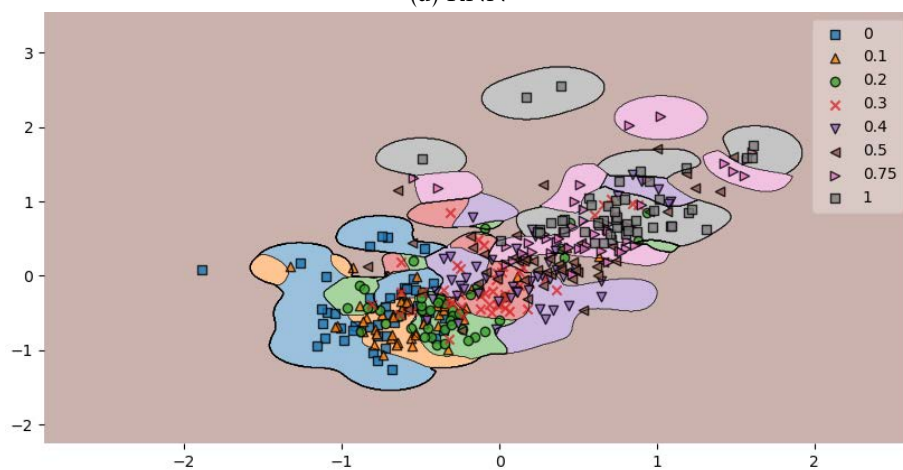


(c) *Random Forest*

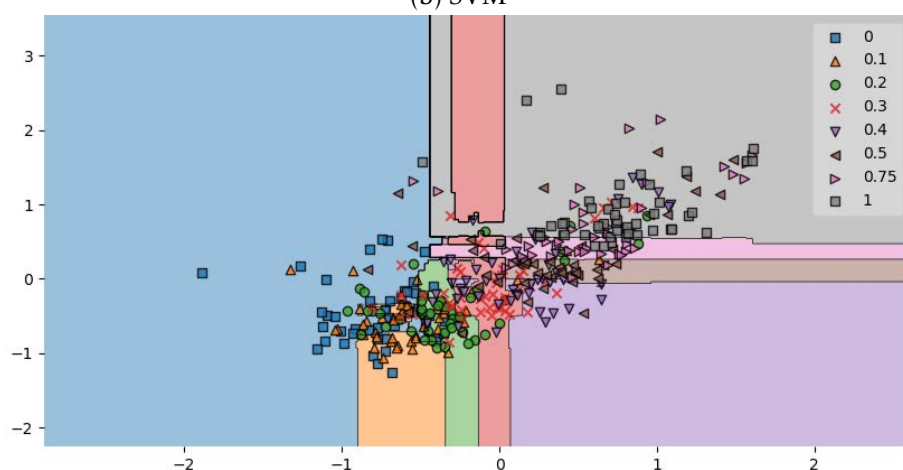
Figura F.84: Regiões de decisão para as concentrações do ácido úrico (mM) dos três algoritmos estudados, baseados no rácio entre os componentes vermelho e azul do espaço RGB e componente H do espaço HSV.



(a) KNN



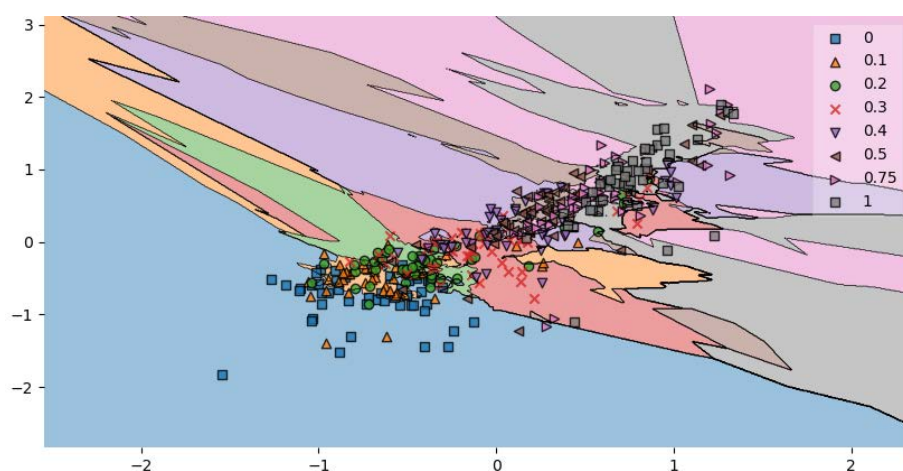
(b) SVM



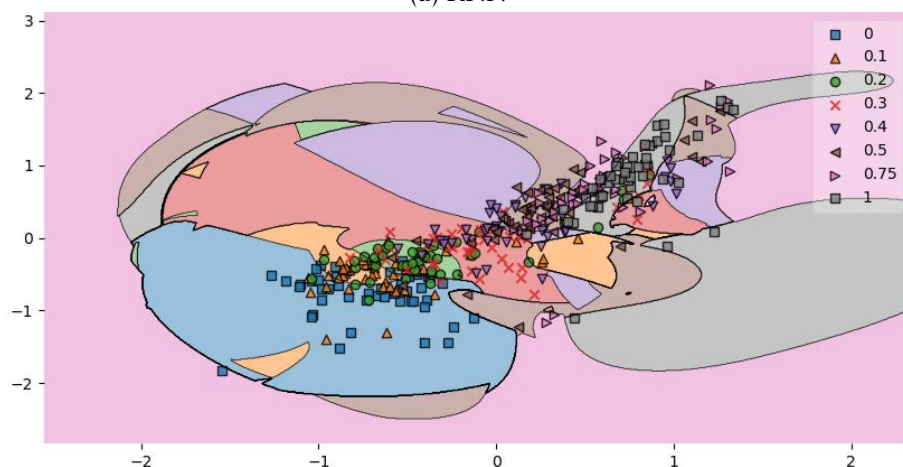
(c) *Random Forest*

Figura F.85: Regiões de decisão para as concentrações do ácido úrico (mM) dos três algoritmos estudados, baseados nos componentes a e b do espaço LAB.

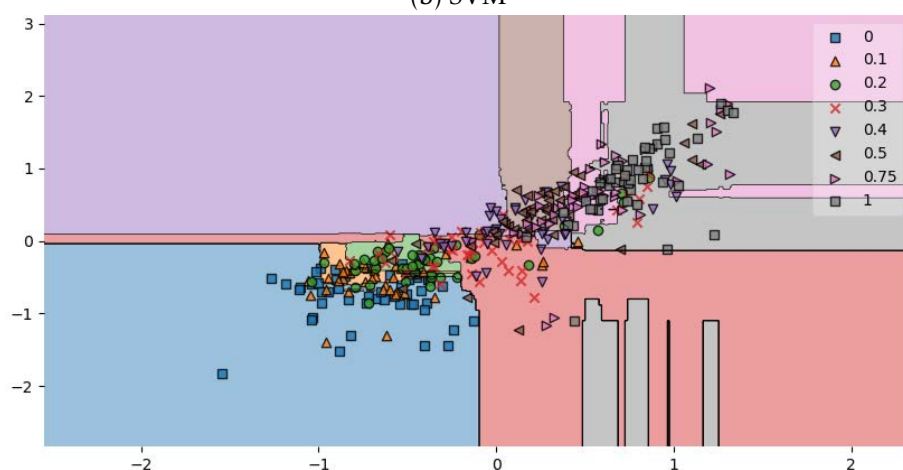




(a) KNN

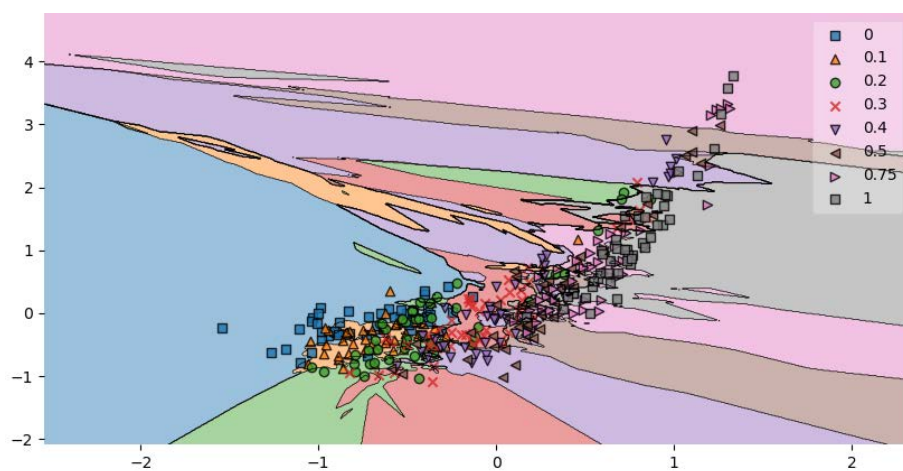


(b) SVM

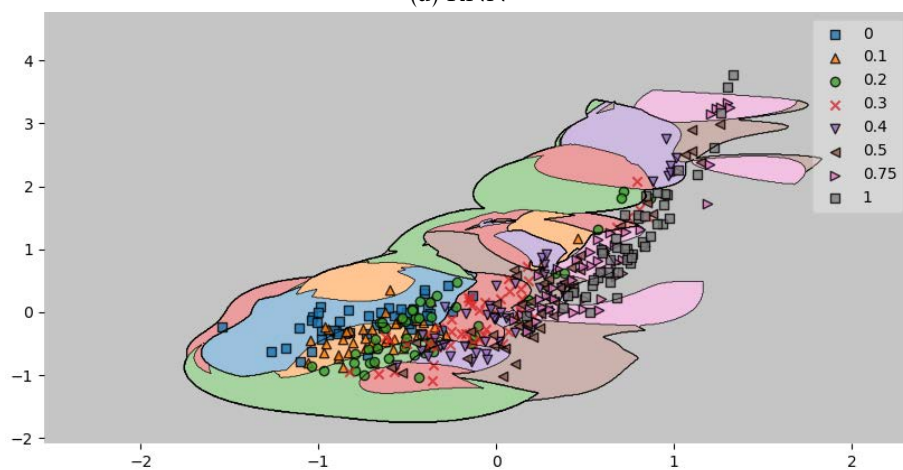


(c) *Random Forest*

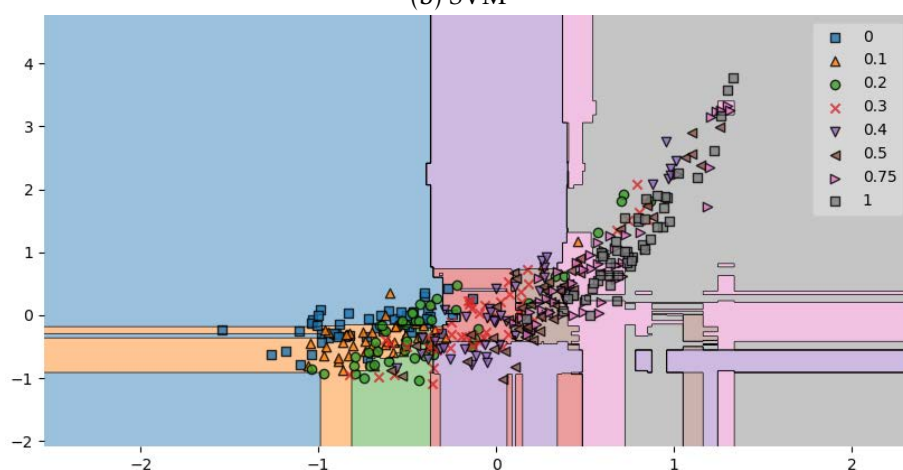
Figura F.86: Regiões de decisão para as concentrações do ácido úrico (mM) dos três algoritmos estudados, baseados nos componentes H e S do espaço HSV.



(a) KNN

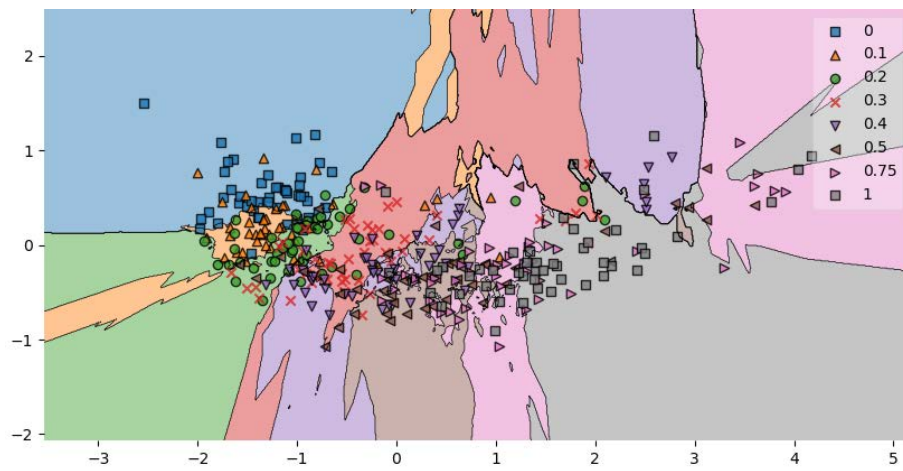


(b) SVM

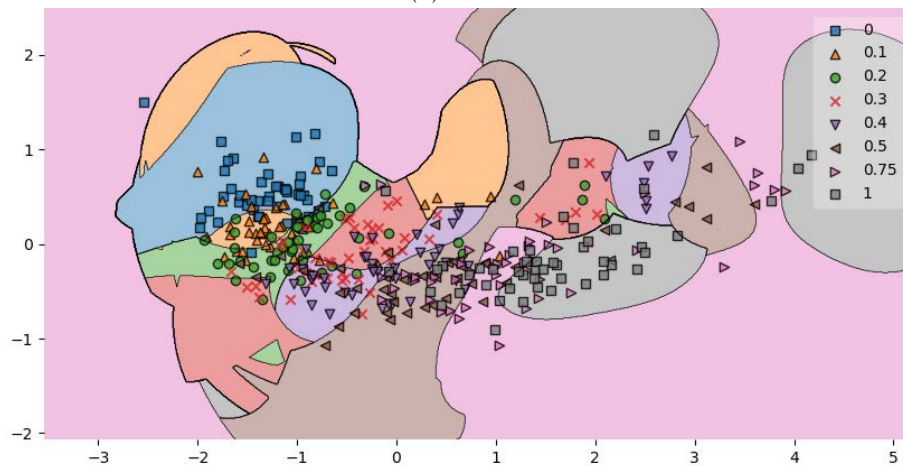


(c) *Random Forest*

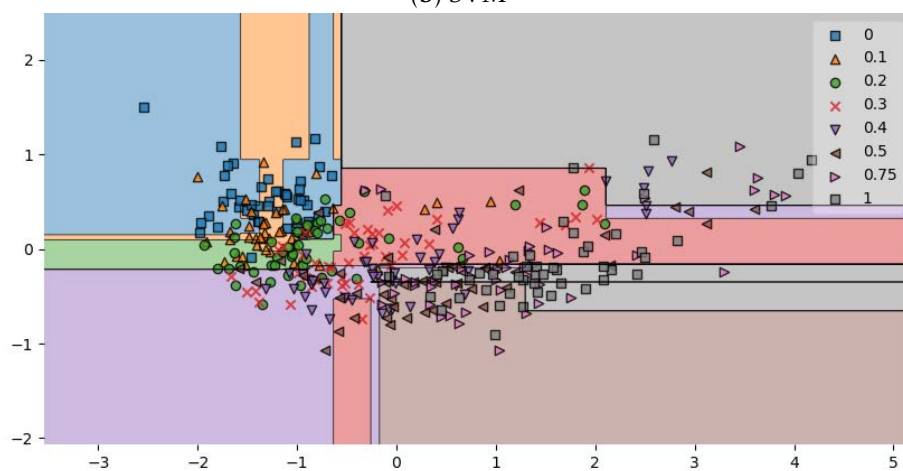
Figura F.87: Regiões de decisão para as concentrações do ácido úrico (mM) dos três algoritmos estudados, baseados nos componentes H e S do espaço HLS.



(a) KNN

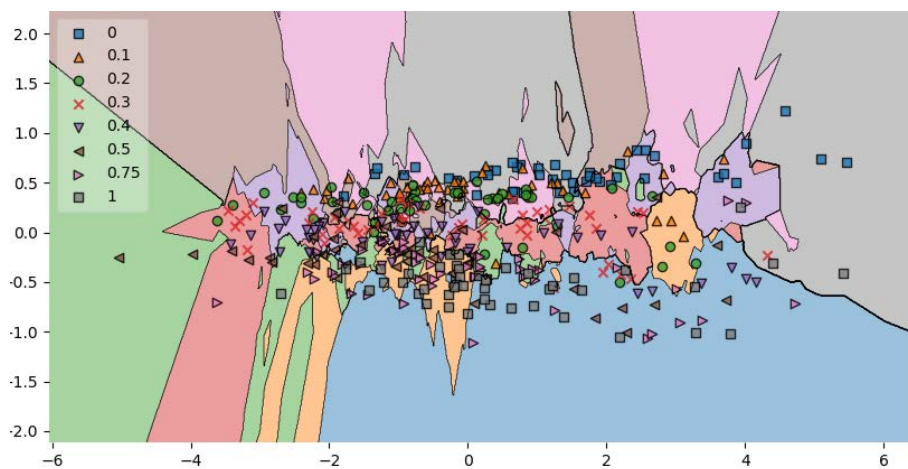


(b) SVM

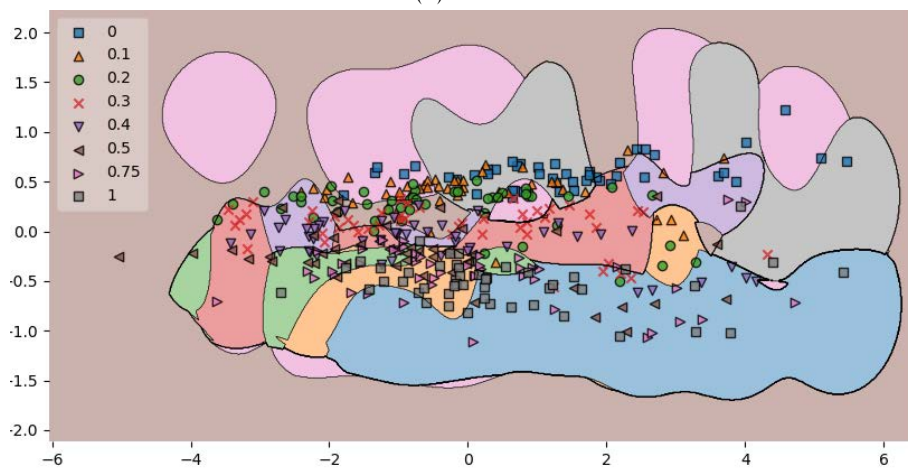


(c) *Random Forest*

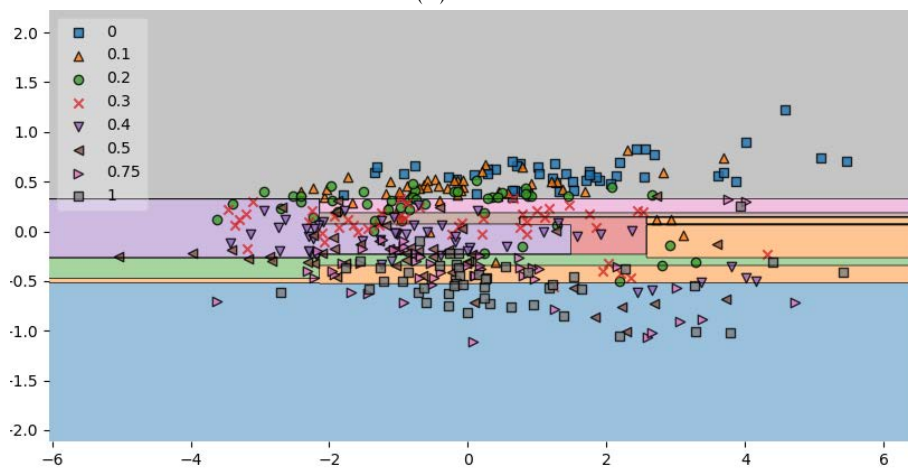
Figura F.88: Regiões de decisão para as concentrações do ácido úrico (mM) dos três algoritmos estudados, baseados nos componentes H e S dos espaços HSV e HLS.



(a) KNN



(b) SVM

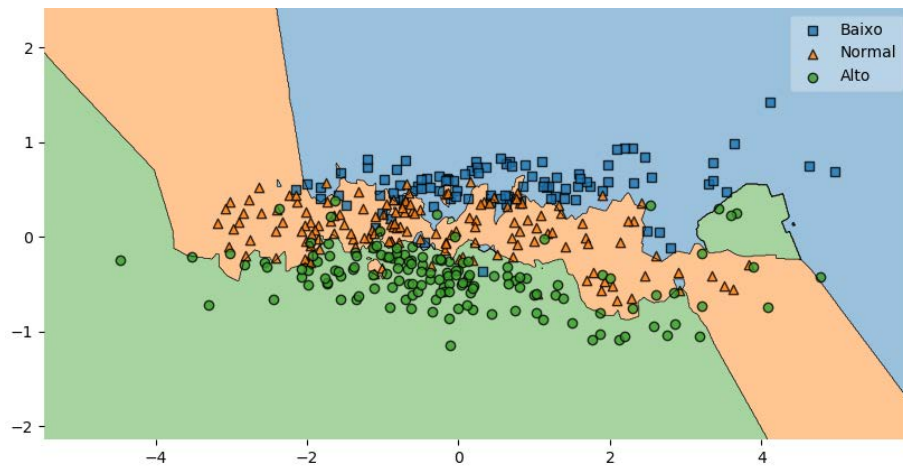


(c) *Random Forest*

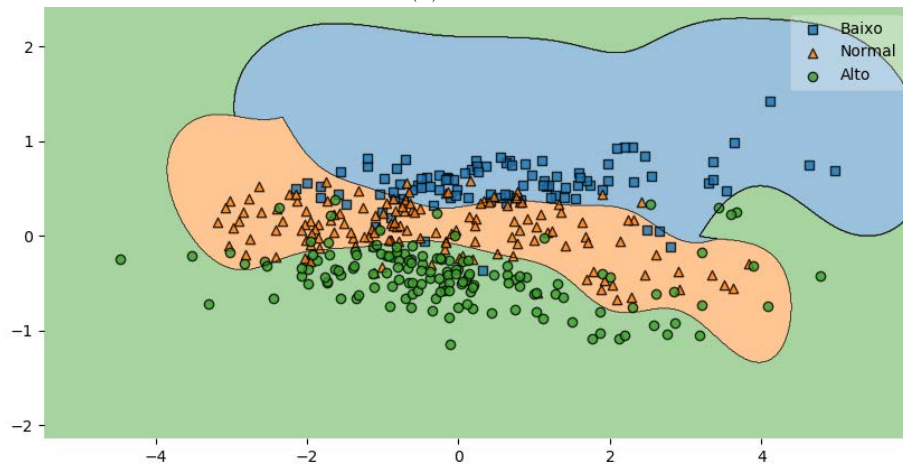
Figura F.89: Regiões de decisão para as concentrações do ácido úrico (mM) dos três algoritmos estudados, baseados na combinação da média dos componentes do modelo RGB com as componentes V do modelo HSV, Z do modelo XYZ, A do modelo LAB e Y do modelo YUV.



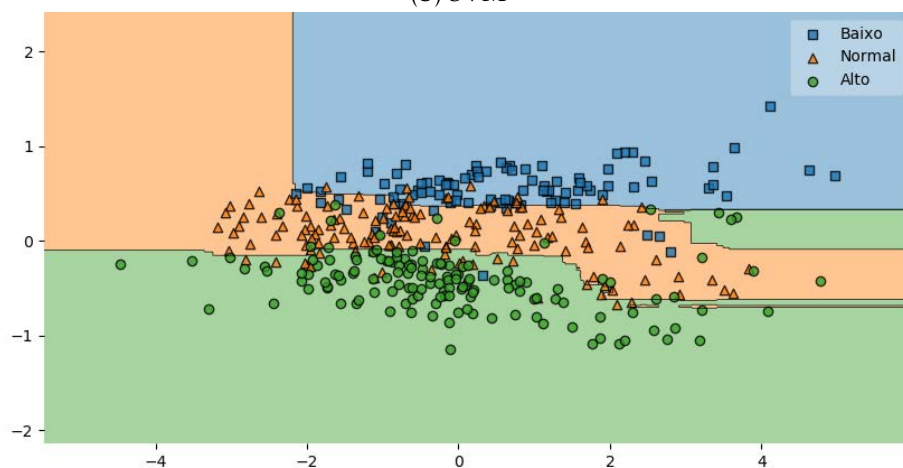
## F.5 Níveis de Ácido Úrico



(a) KNN

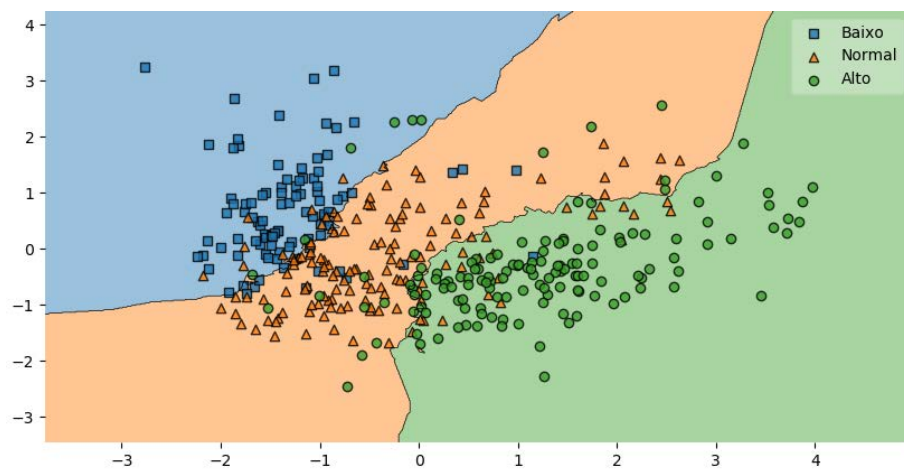


(b) SVM

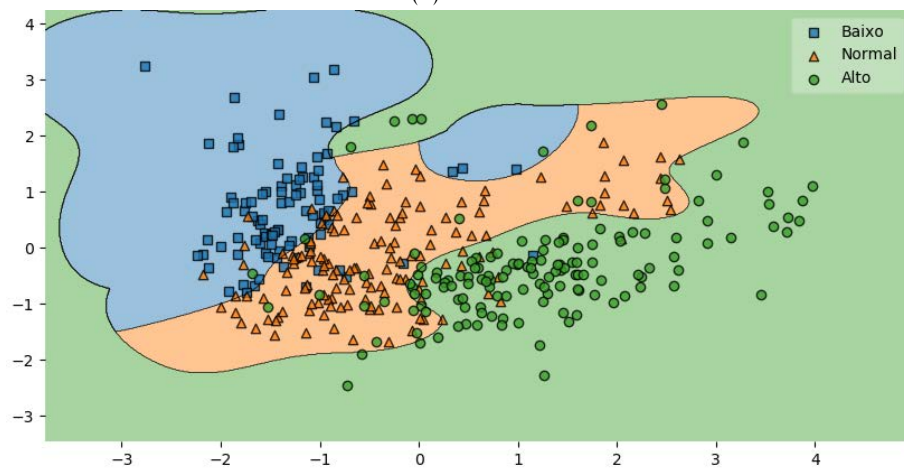


(c) Random Forest

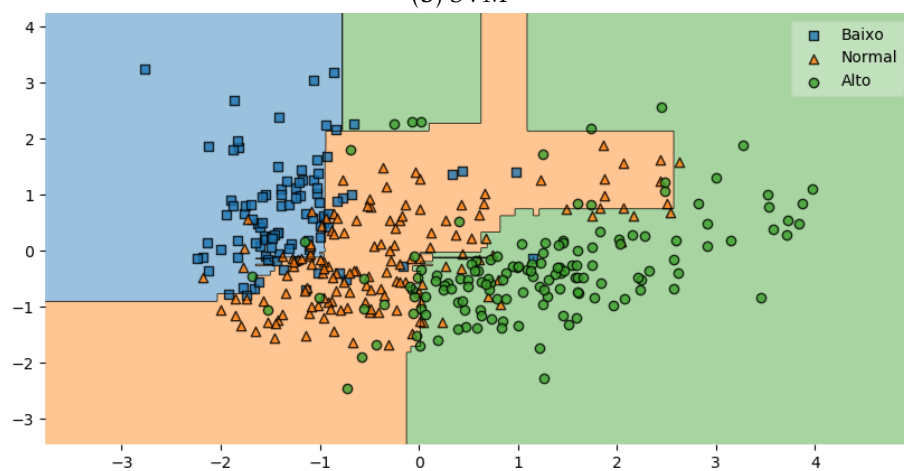
Figura F.90: Regiões de decisão para os níveis de referência do ácido úrico dos três algoritmos estudados, baseados na média dos componentes RGB.



(a) KNN

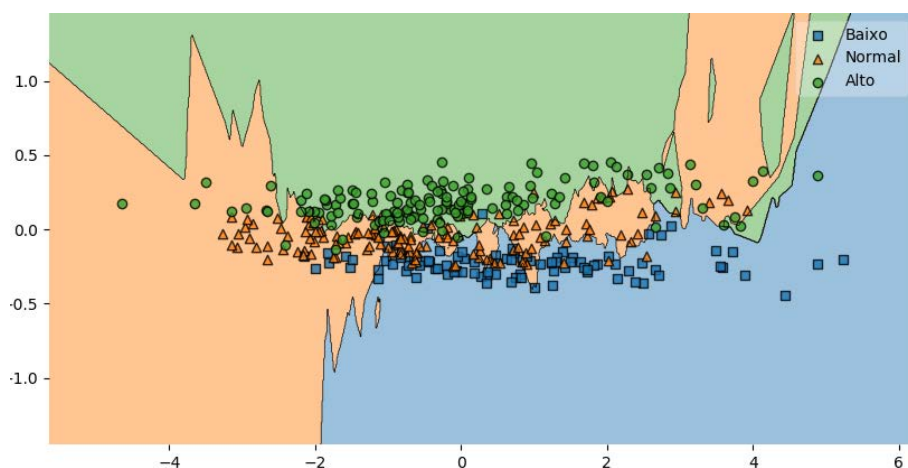


(b) SVM

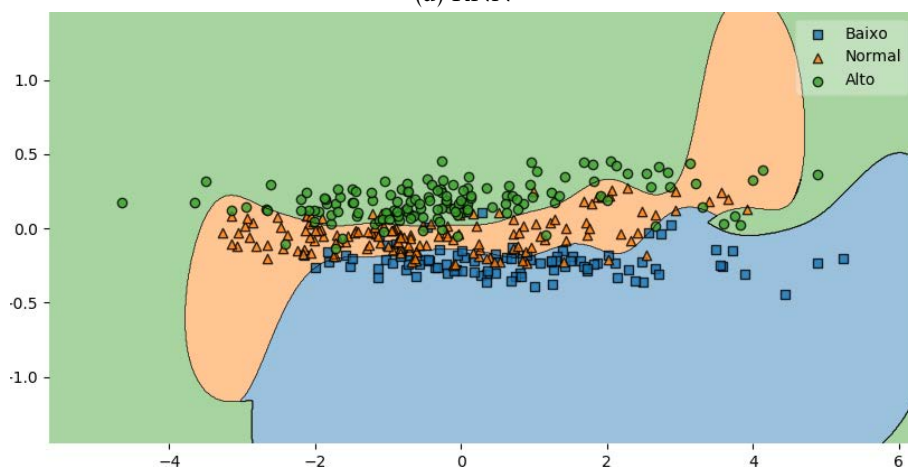


(c) *Random Forest*

Figura F.91: Regiões de decisão para os níveis de referência do ácido úrico dos três algoritmos estudados, baseados na média dos componentes HSV.



(a) KNN



(b) SVM

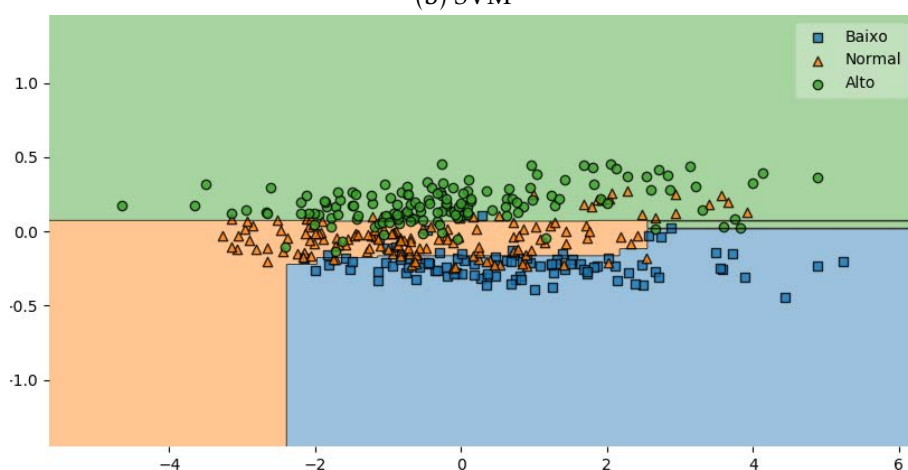
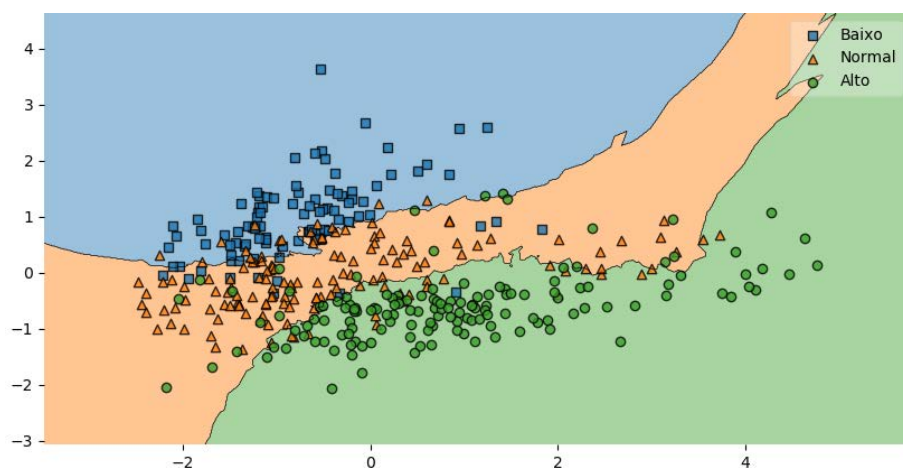
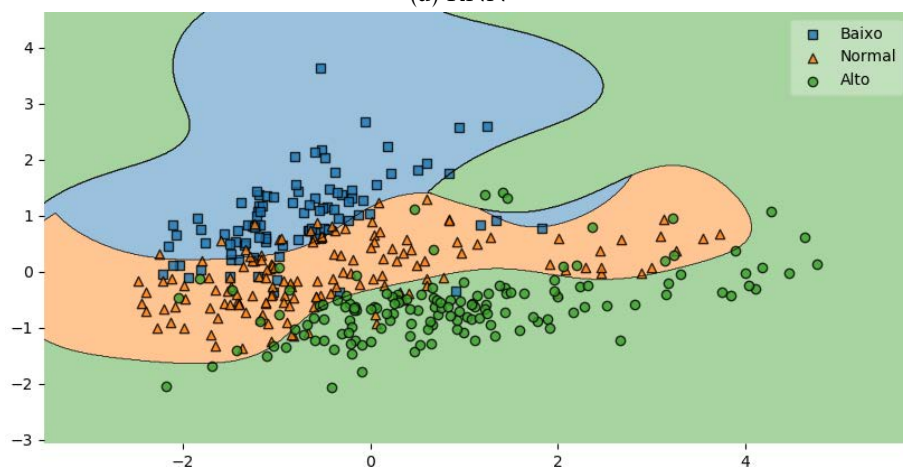
(c) *Random Forest*

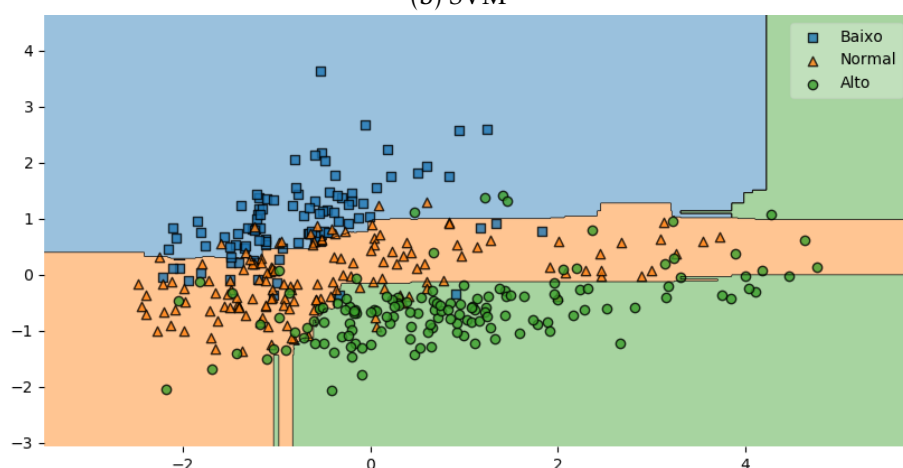
Figura F.92: Regiões de decisão para os níveis de referência do ácido úrico dos três algoritmos estudados, baseados na média dos componentes XYZ.



(a) KNN



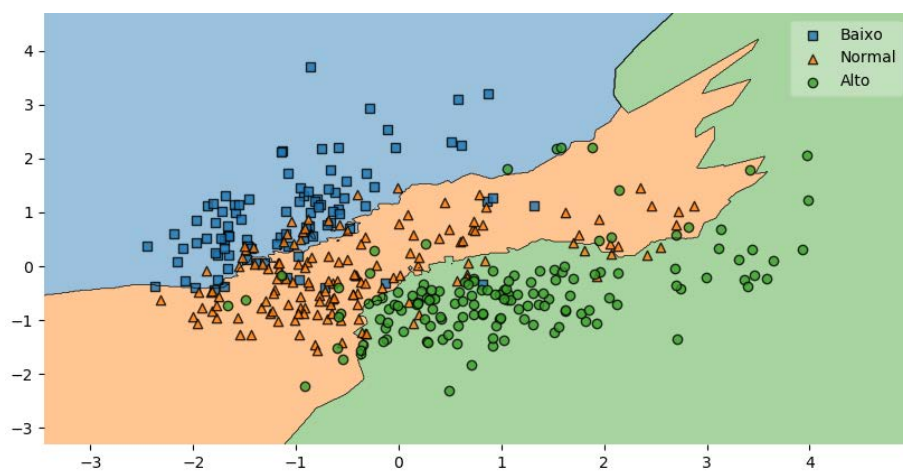
(b) SVM



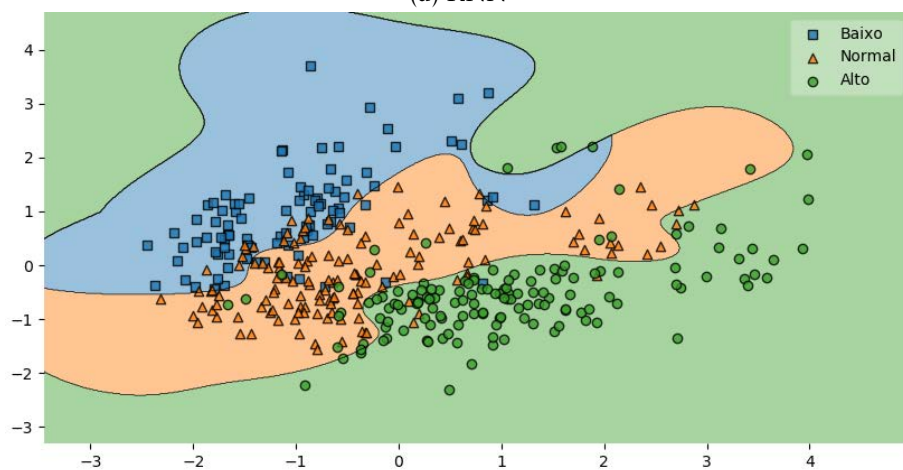
(c) *Random Forest*

Figura F.93: Regiões de decisão para os níveis de referência do ácido úrico dos três algoritmos estudados, baseados na média dos componentes HLS.





(a) KNN



(b) SVM

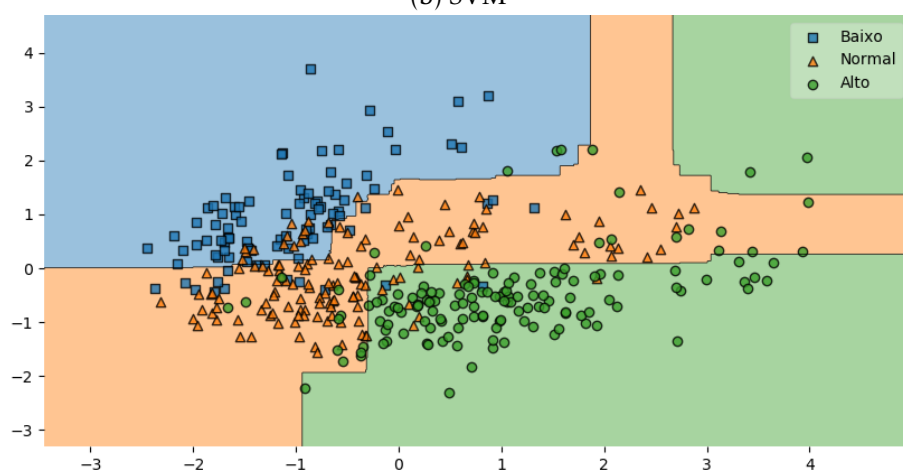
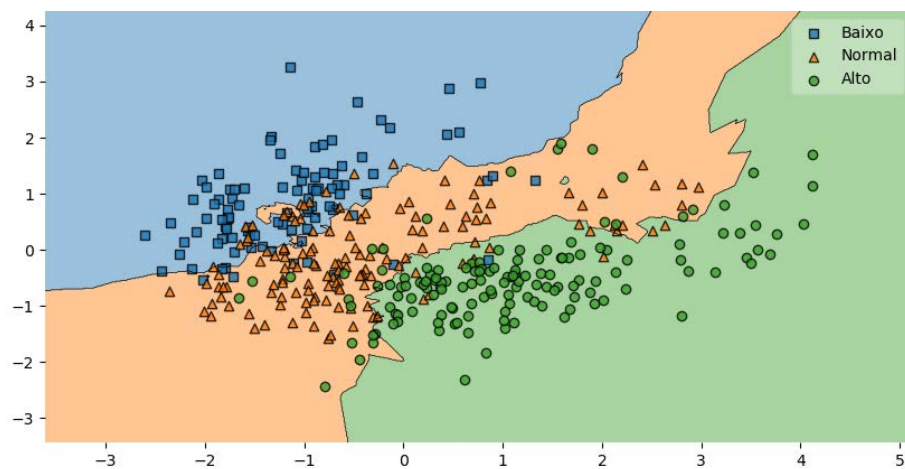
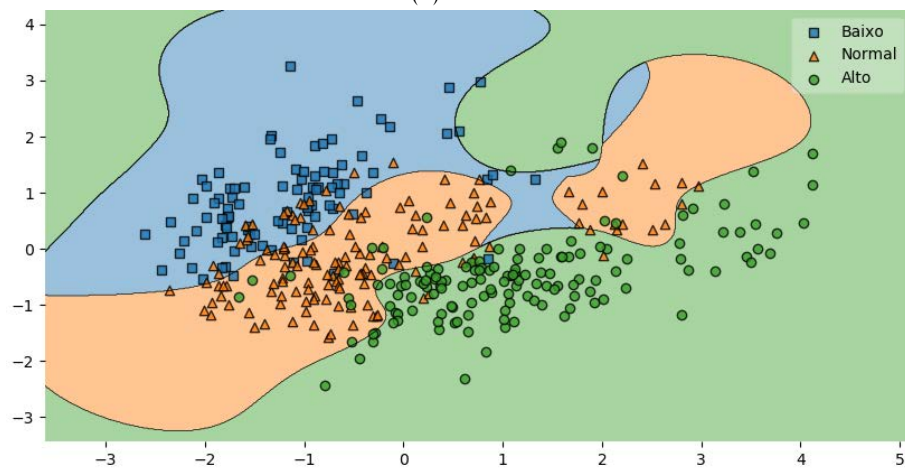
(c) *Random Forest*

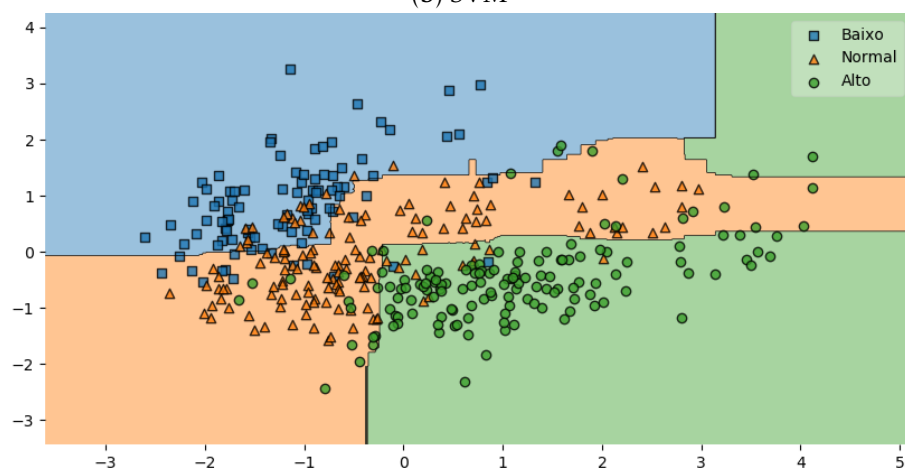
Figura F.94: Regiões de decisão para os níveis de referência do ácido úrico dos três algoritmos estudados, baseados na média dos componentes LAB.



(a) KNN

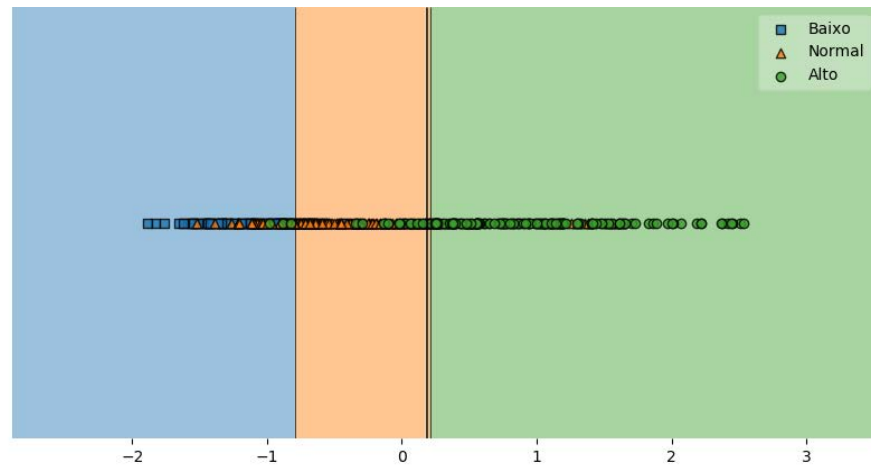


(b) SVM

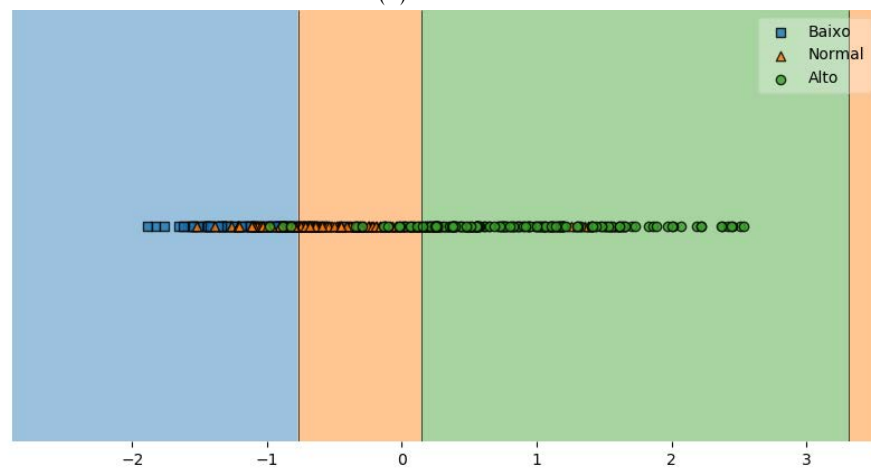


(c) *Random Forest*

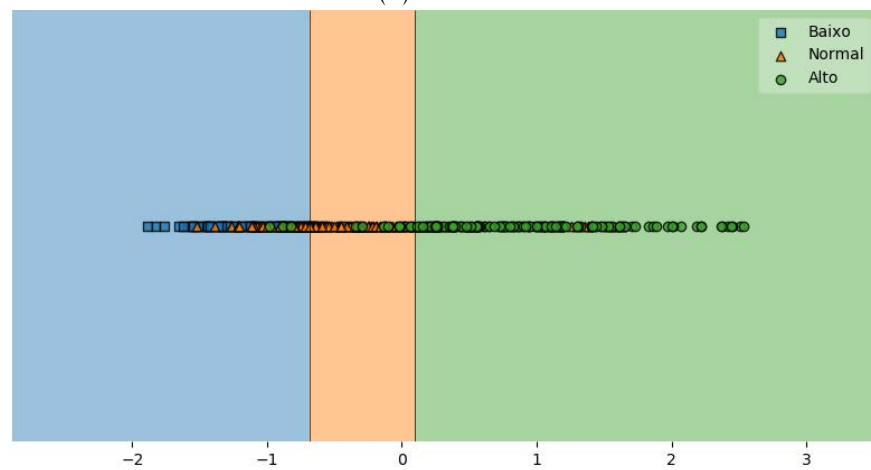
Figura F.95: Regiões de decisão para os níveis de referência do ácido úrico dos três algoritmos estudados, baseados na média dos componentes YUV.



(a) KNN

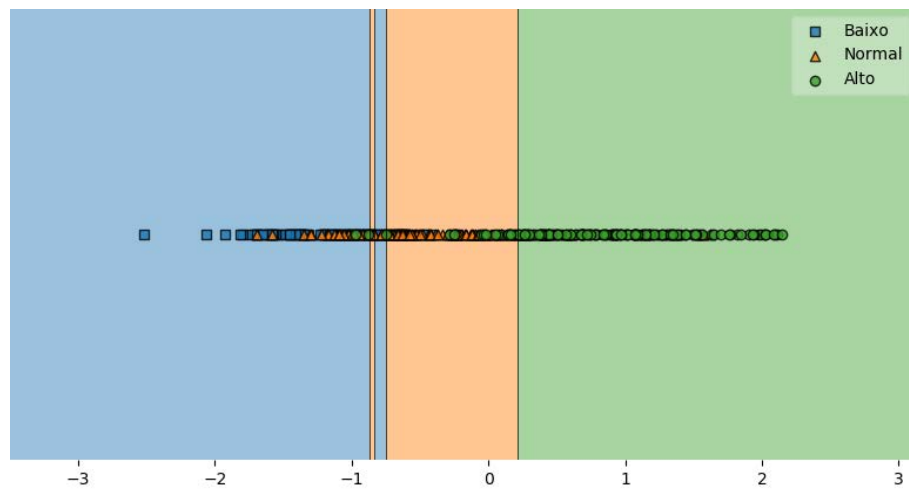


(b) SVM

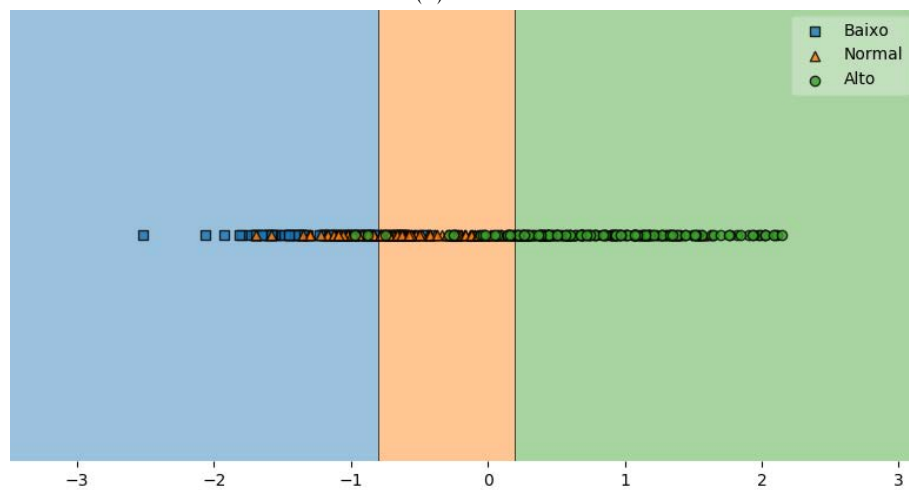


(c) Random Forest

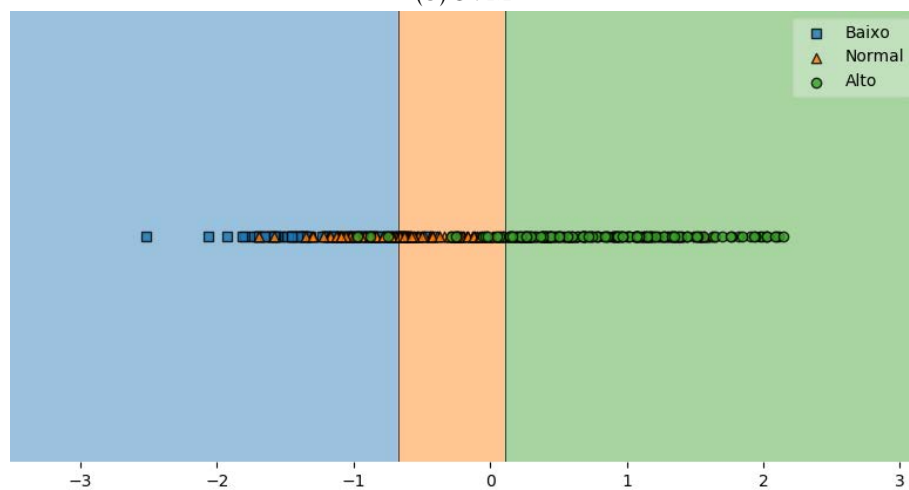
Figura F.96: Regiões de decisão para os níveis de referência do ácido úrico dos três algoritmos estudados, baseados no rácio entre os componentes vermelho e azul do espaço RGB.



(a) KNN

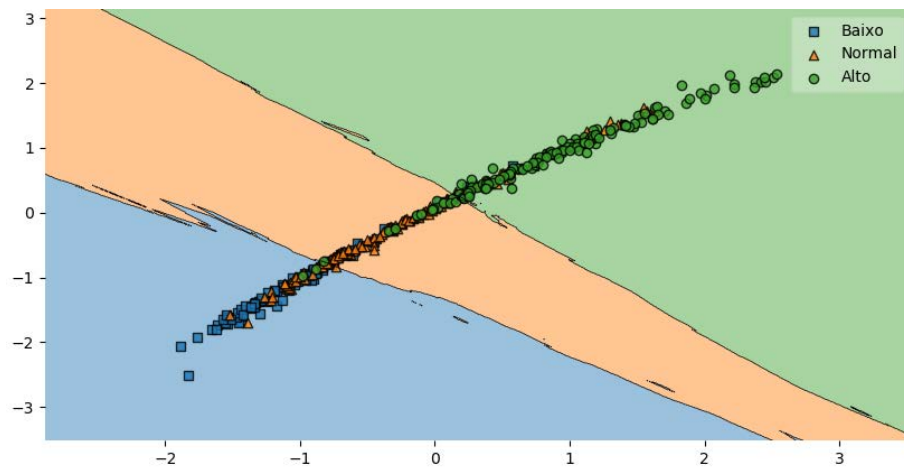


(b) SVM

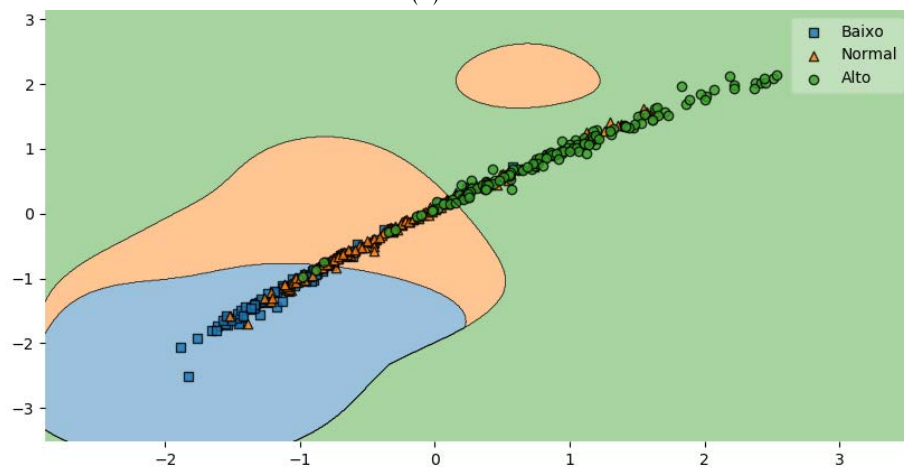


(c) Random Forest

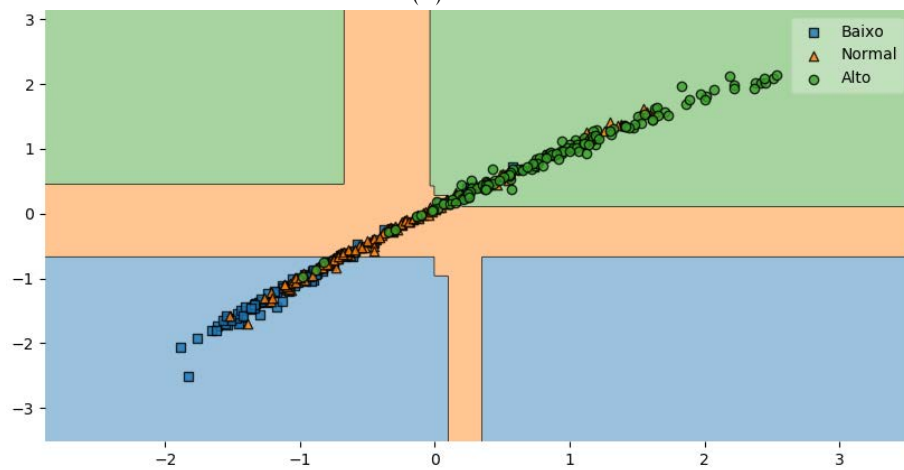
Figura F.97: Regiões de decisão para os níveis de referência do ácido úrico dos três algoritmos estudados, baseados no componente H do espaço HSV.



(a) KNN

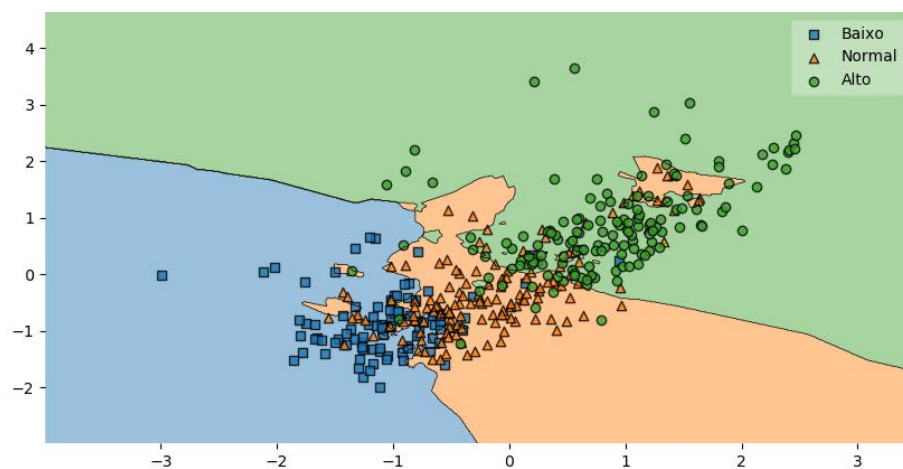


(b) SVM

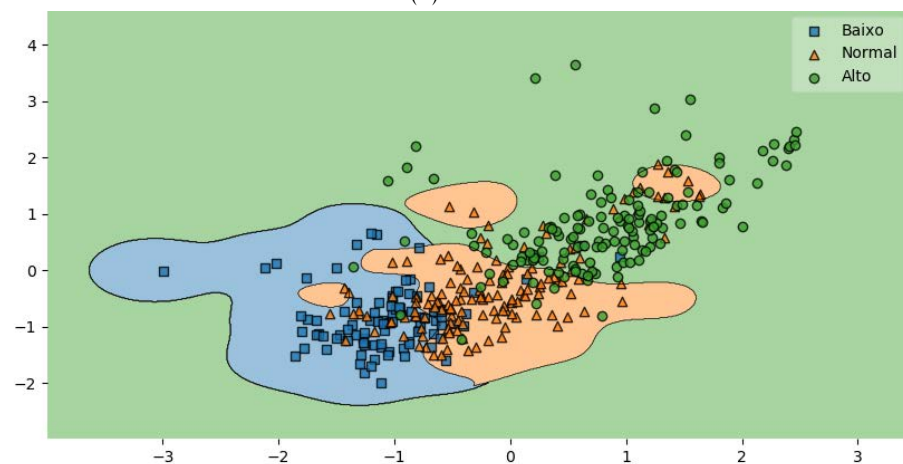


(c) Random Forest

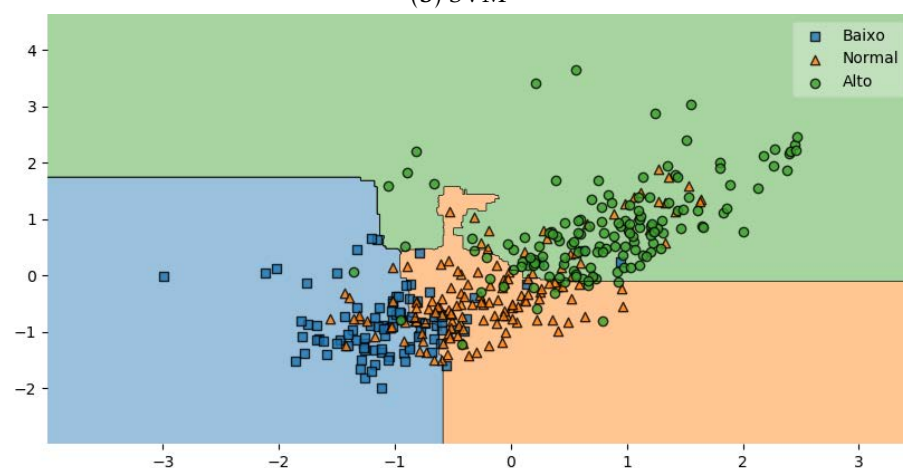
Figura F.98: Regiões de decisão para os níveis de referência do ácido úrico dos três algoritmos estudados, baseados no rácio entre os componentes vermelho e azul do espaço RGB e componente H do espaço HSV.



(a) KNN

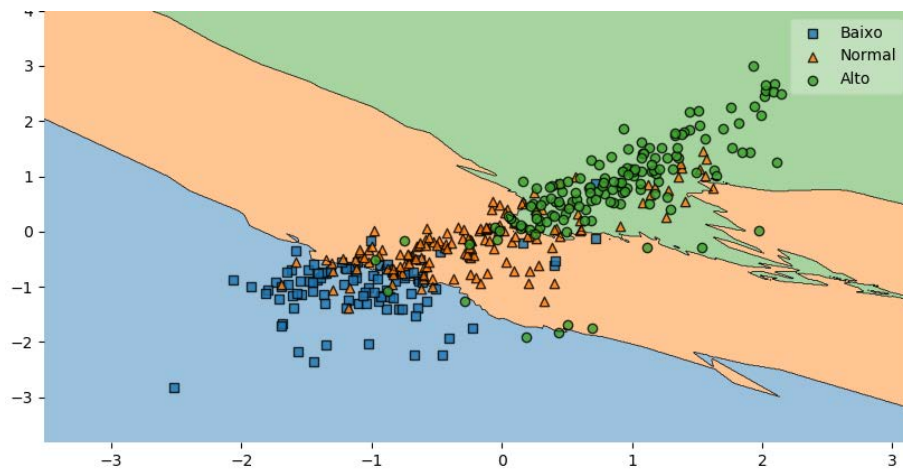


(b) SVM

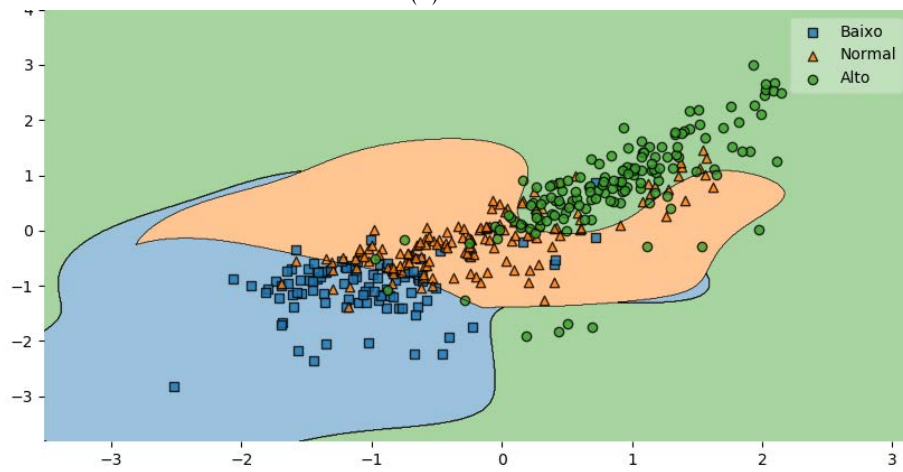


(c) *Random Forest*

Figura F.99: Regiões de decisão para os níveis de referência do ácido úrico dos três algoritmos estudados, baseados nos componentes a e b do espaço LAB.



(a) KNN



(b) SVM

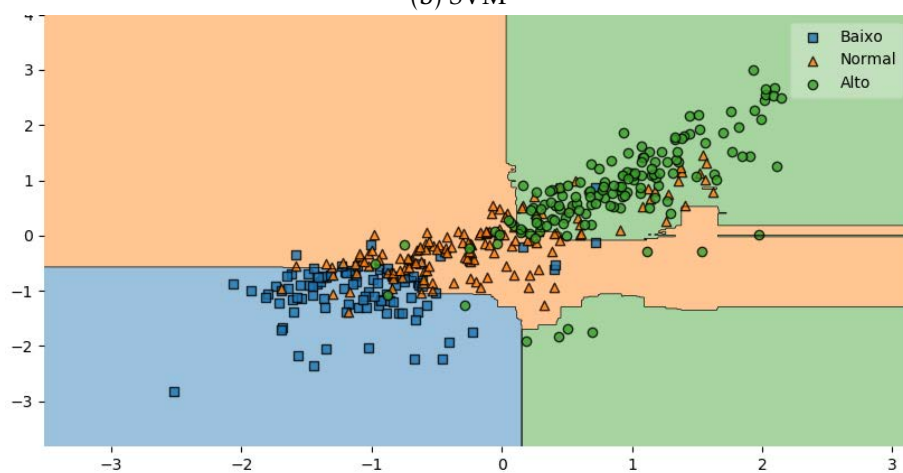
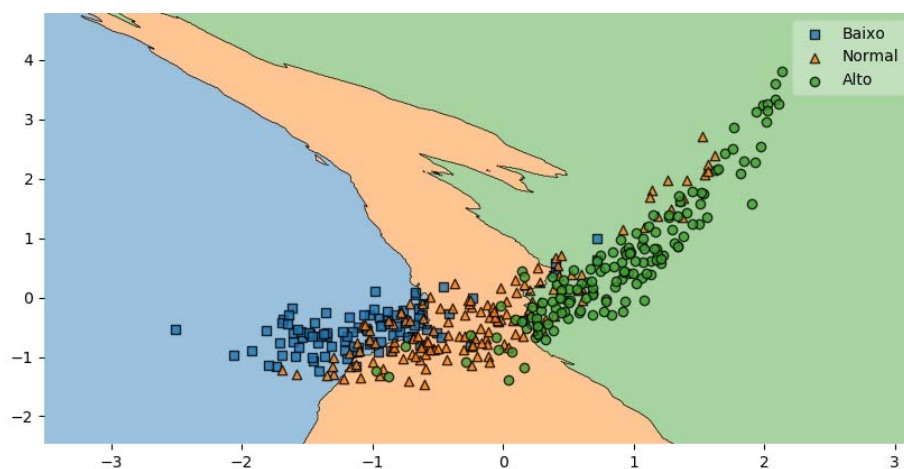
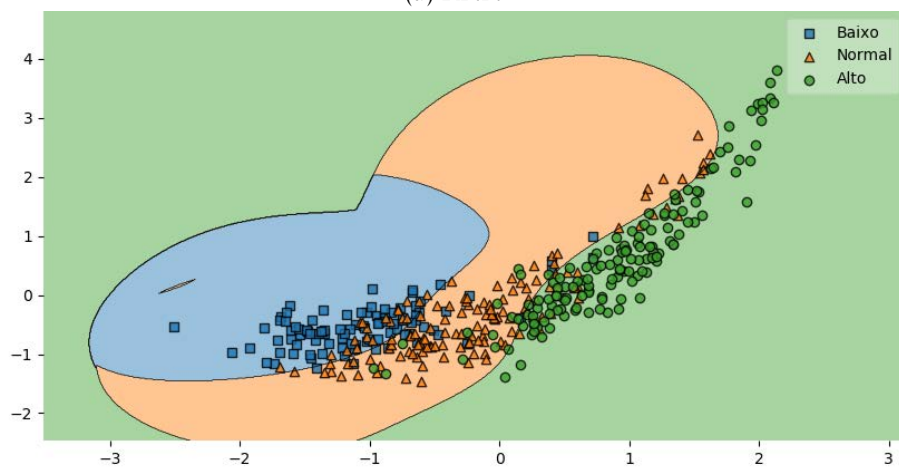
(c) *Random Forest*

Figura F.100: Regiões de decisão para os níveis de referência do ácido úrico dos três algoritmos estudados, baseados nos componentes H e S do espaço HSV.

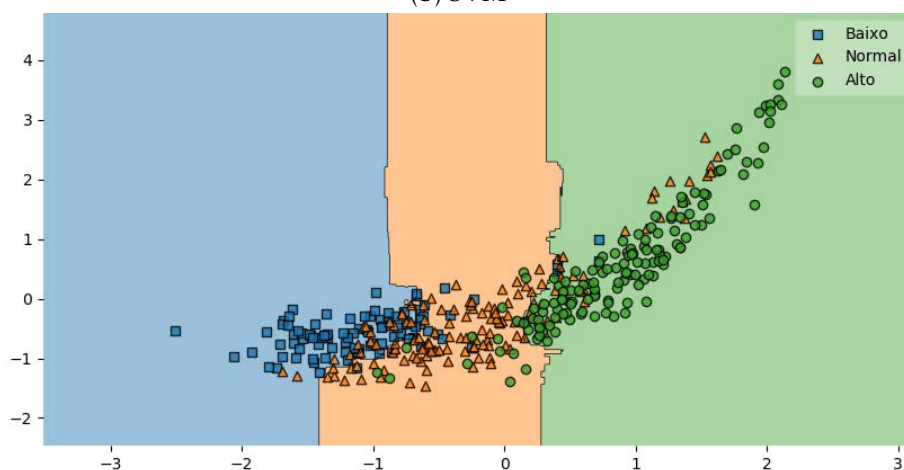




(a) KNN



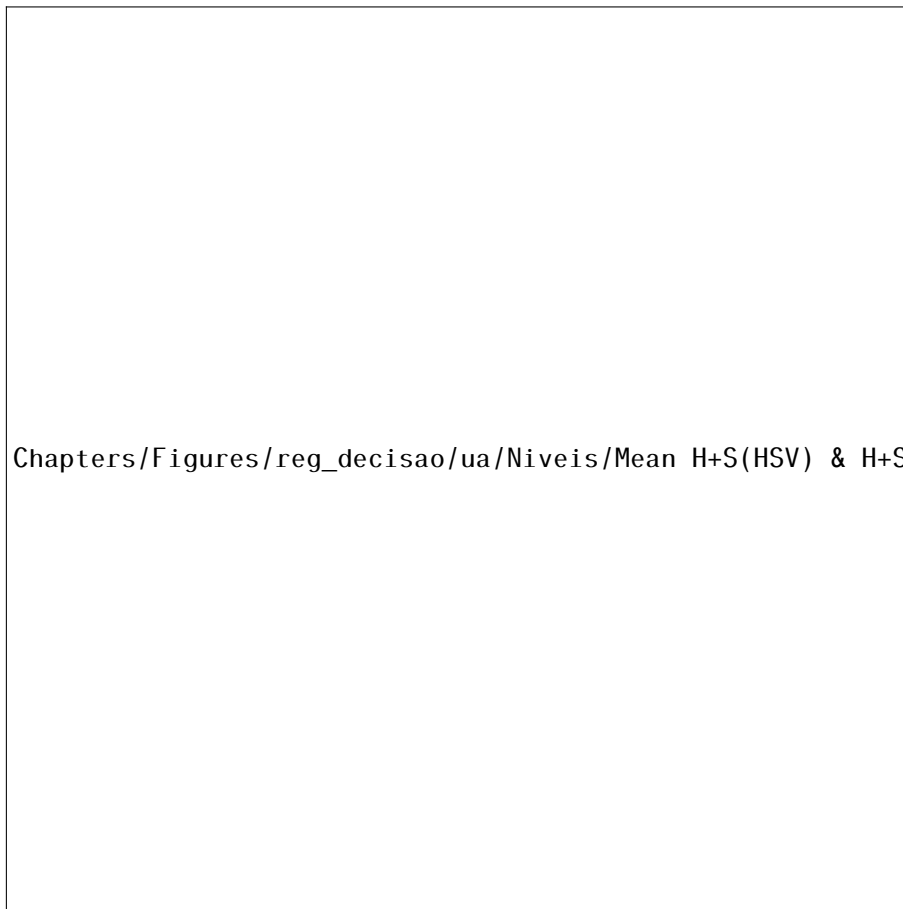
(b) SVM



(c) *Random Forest*

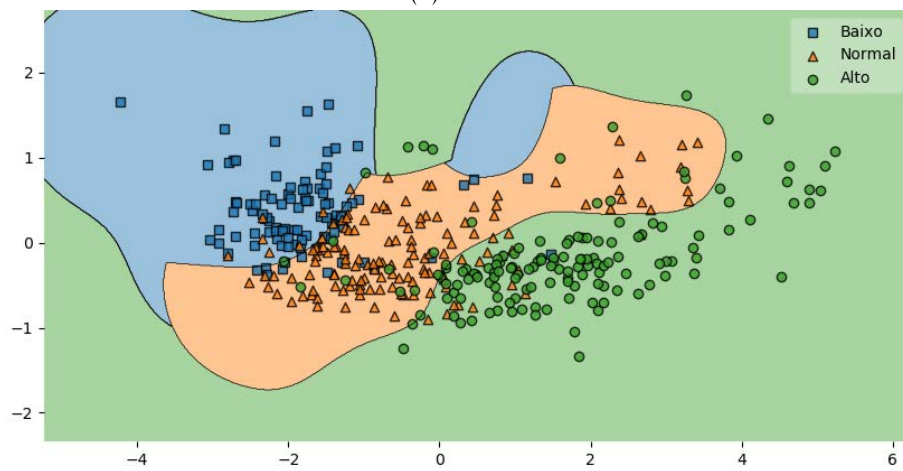
Figura F.101: Regiões de decisão para os níveis de referência do ácido úrico dos três algoritmos estudados, baseados nos componentes H e S do espaço HLS.



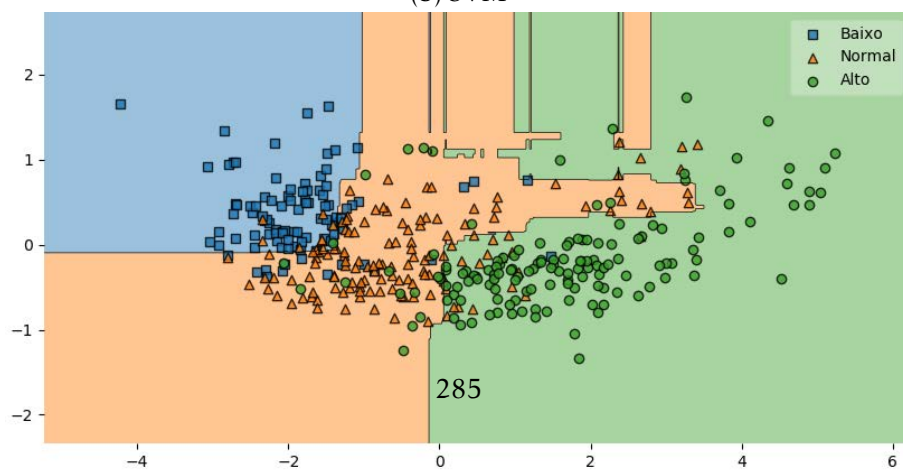


Chapters/Figures/reg\_decisao/ua/Niveis/Mean H+S(HSV) & H+S(HLS) basic non pca knn

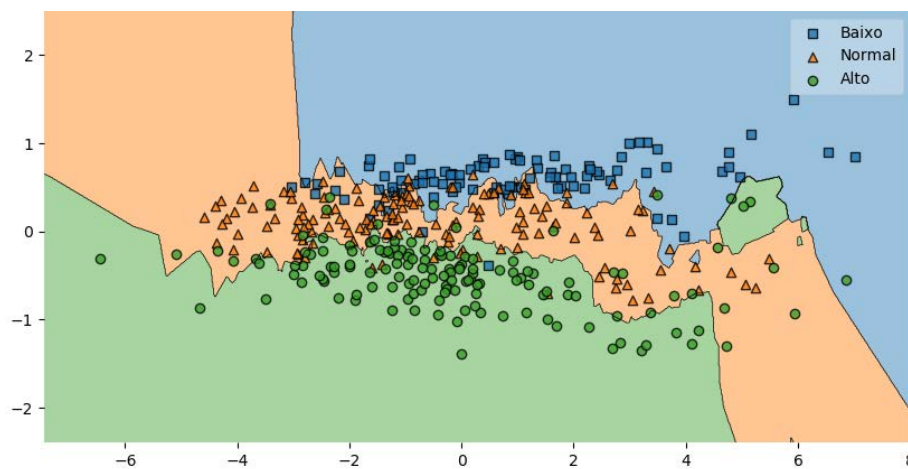
(a) KNN



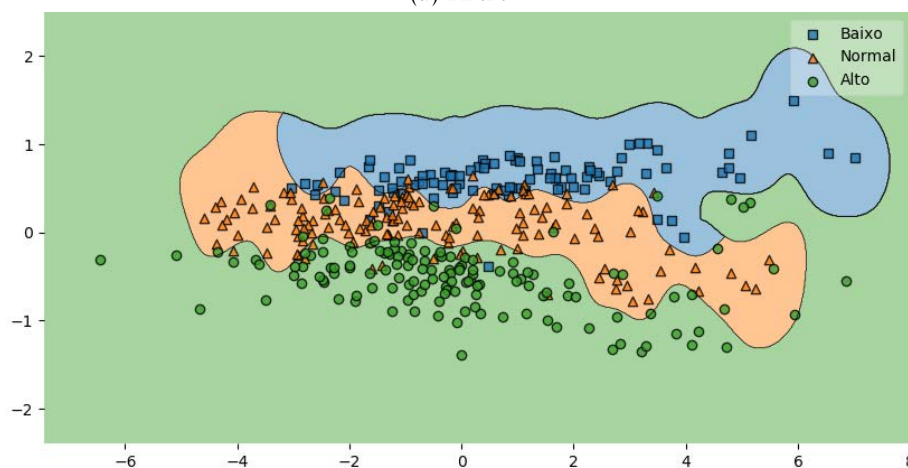
(b) SVM



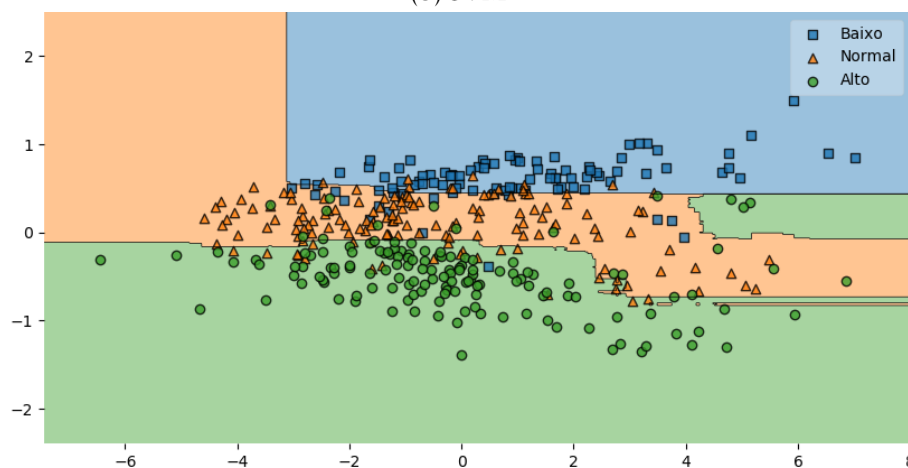
(c) Random Forest



(a) KNN



(b) SVM



(c) Random Forest

Figura F.103: Regiões de decisão para os níveis de referência do ácido úrico dos três algoritmos estudados, baseados na combinação da média dos componentes do modelo RGB com as componentes V do modelo HSV, Z do modelo XYZ, A do modelo LAB e Y do modelo YUV.

## RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

Neste apêndice são apresentadas as visualizações dos conjuntos de dados do ácido úrico e da glucose, obtidas através da aplicação do PCA.

### G.1 Concentrações de Glucose

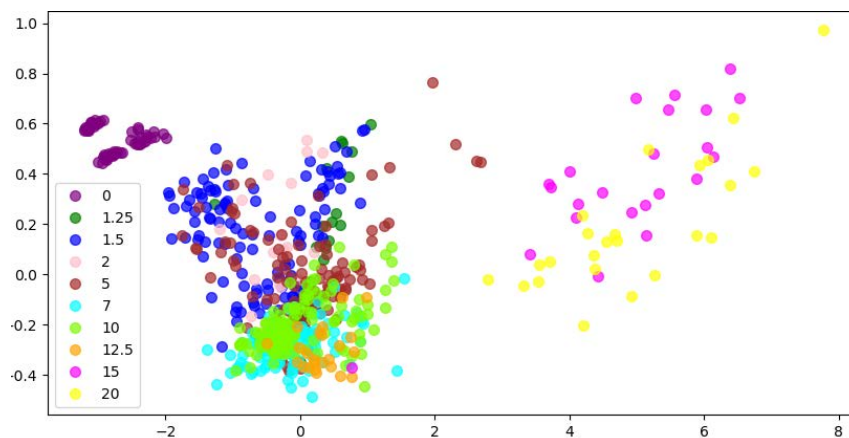


Figura G.1: Representação através de PCA das concentrações de glucose (mM), baseada na média dos componentes RGB.

APÊNDICE G. RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

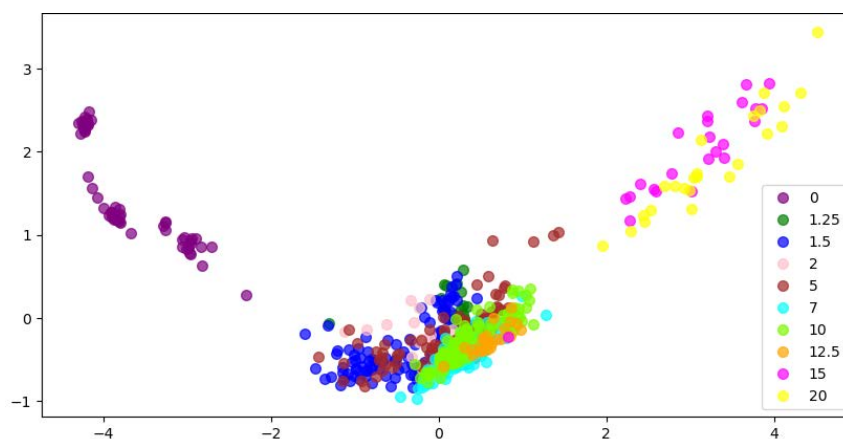


Figura G.2: Representação através de PCA das concentrações de glucose (mM), baseada na média dos componentes HSV.

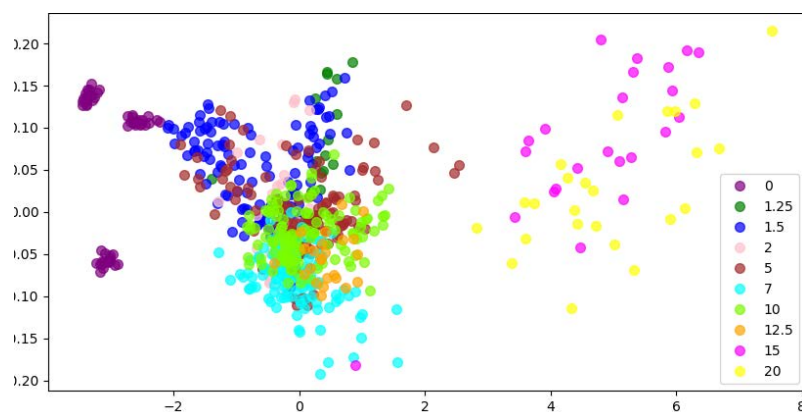


Figura G.3: Representação através de PCA das concentrações de glucose (mM), baseada na média dos componentes XYZ.

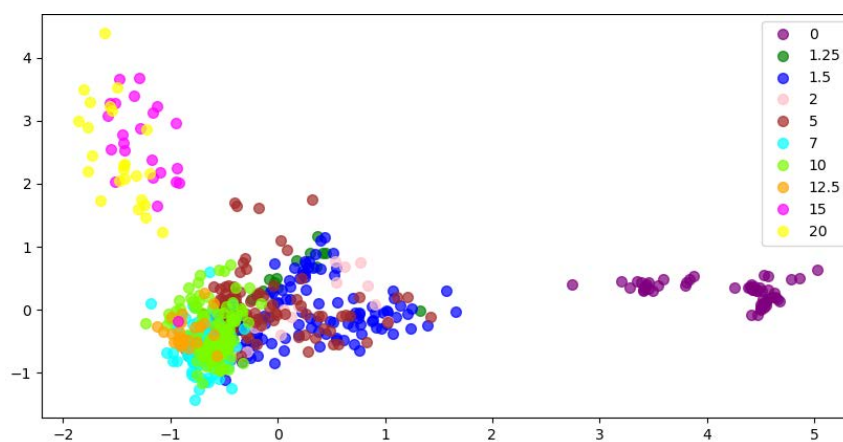


Figura G.4: Representação através de PCA das concentrações de glucose (mM), baseada na média dos componentes HLS.

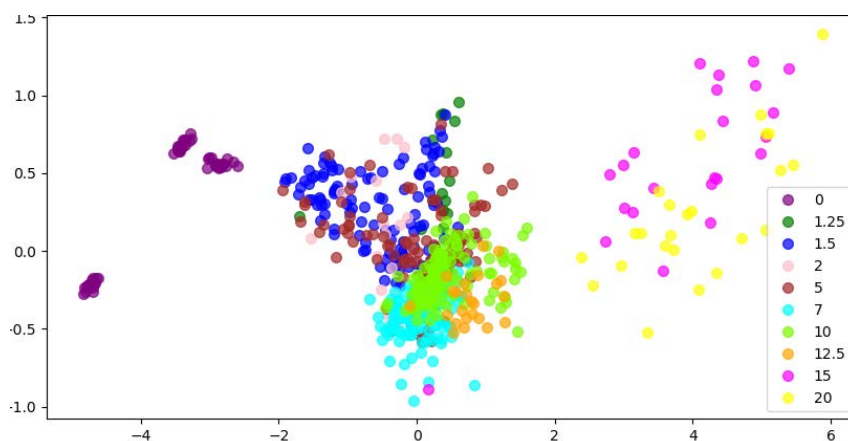


Figura G.5: Representação através de PCA das concentrações de glucose (mM), baseada na média dos componentes LAB.

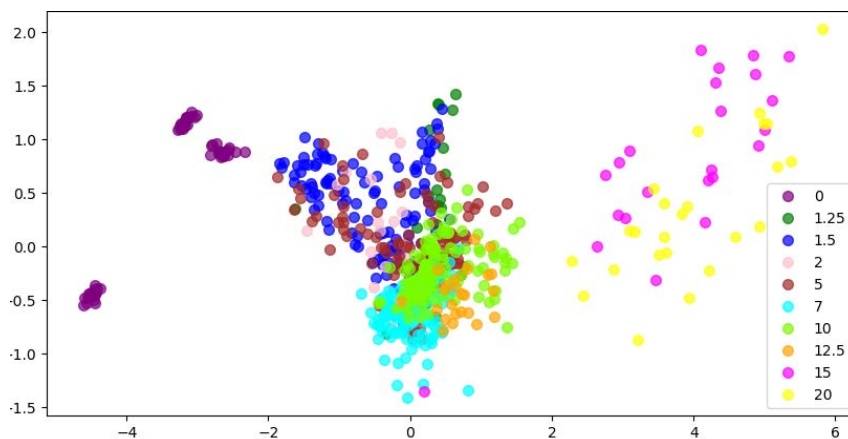


Figura G.6: Representação através de PCA das concentrações de glucose (mM), baseada na média dos componentes YUV.

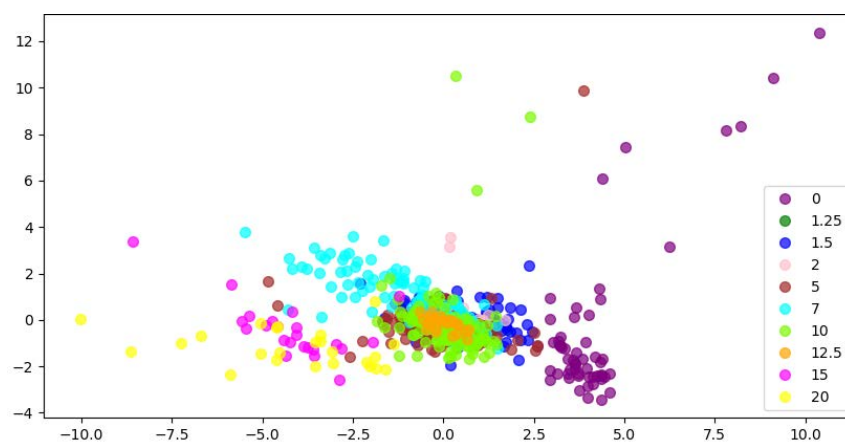


Figura G.7: Representação através de PCA das concentrações de glucose (mM), baseada na média, desvio padrão e assimetria dos componentes RGB.

APÊNDICE G. RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

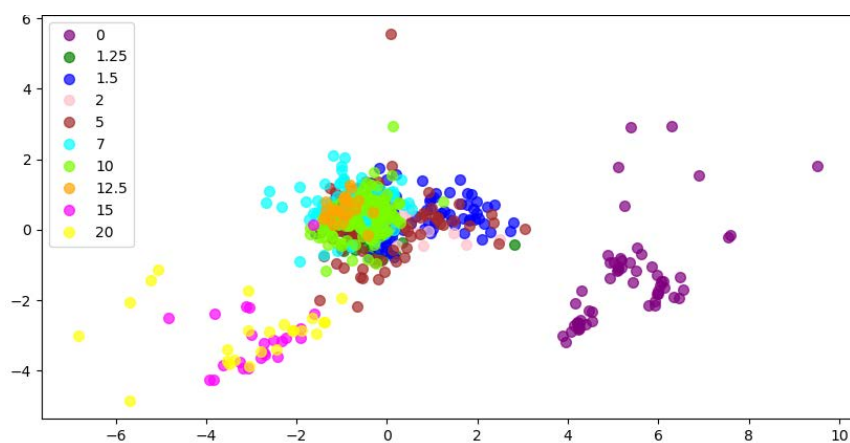


Figura G.8: Representação através de PCA das concentrações de glucose (mM), baseada na média, desvio padrão e assimetria dos componentes HSV.

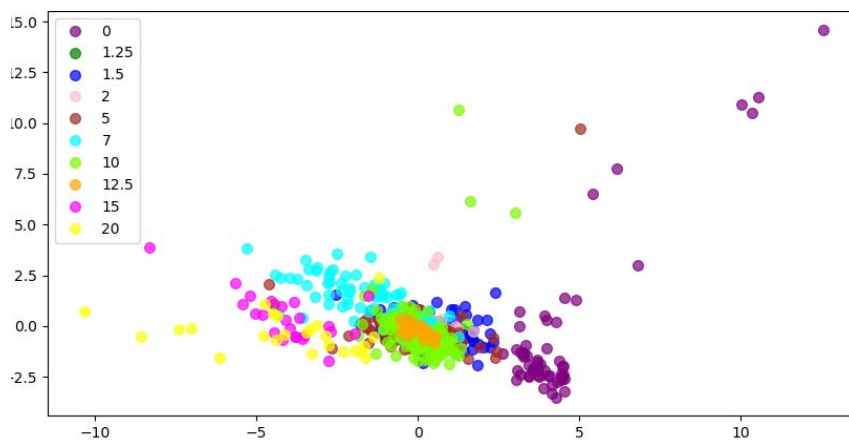


Figura G.9: Representação através de PCA das concentrações de glucose (mM), baseada na média, desvio padrão e assimetria dos componentes XYZ.

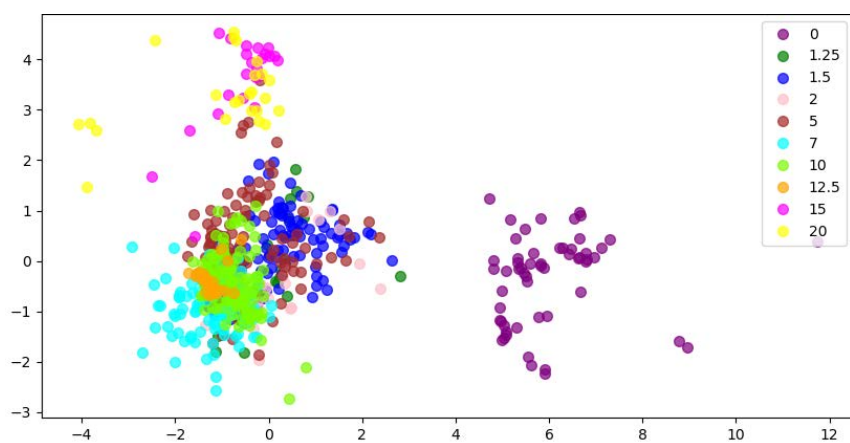


Figura G.10: Representação através de PCA das concentrações de glucose (mM), baseada na média, desvio padrão e assimetria dos componentes HLS.

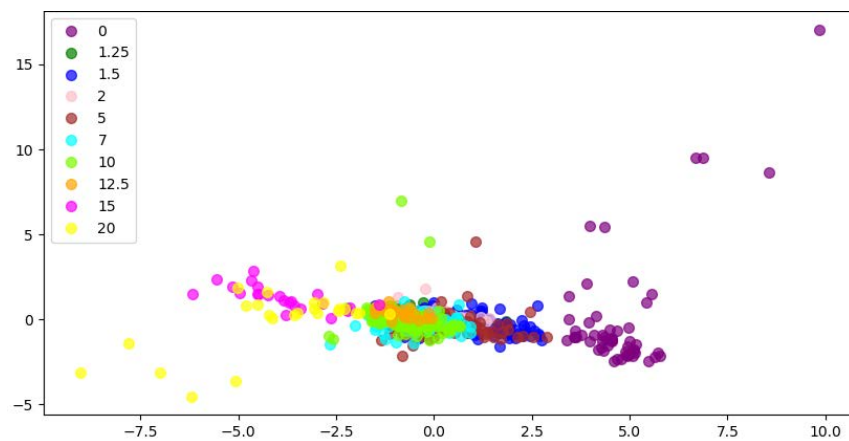


Figura G.11: Representação através de PCA das concentrações de glucose (mM), baseada na média, desvio padrão e assimetria dos componentes LAB.

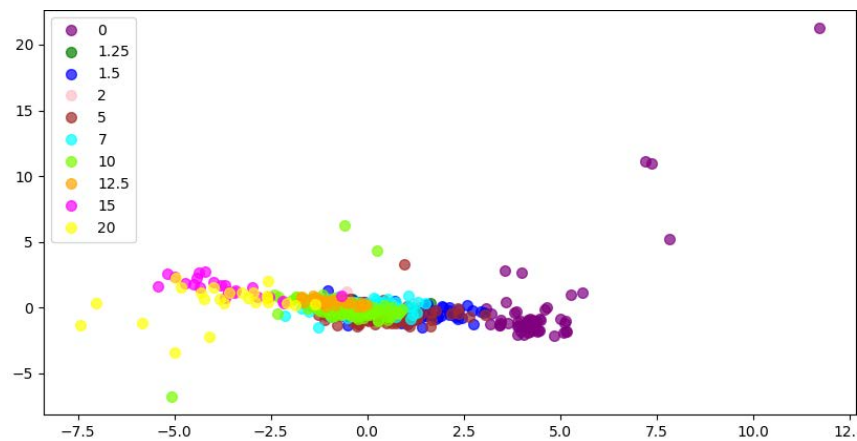


Figura G.12: Representação através de PCA das concentrações de glucose (mM), baseada na média, desvio padrão e assimetria dos componentes YUV.

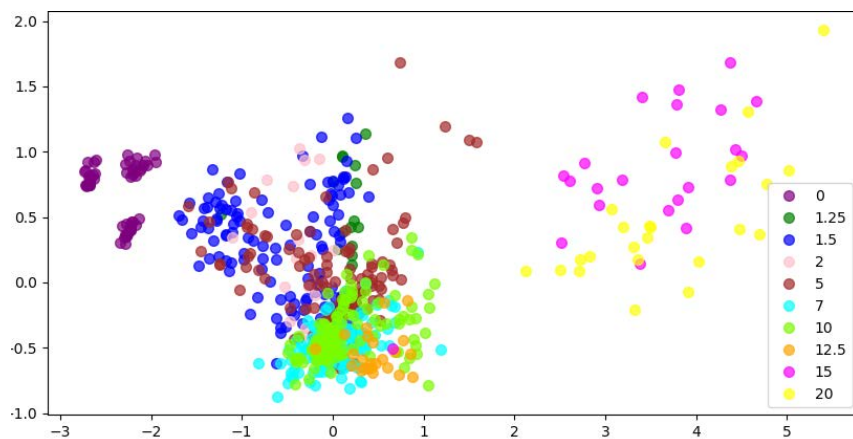


Figura G.13: Representação através de PCA das concentrações de glucose (mM), baseada nos componentes S e V do espaço HSV.

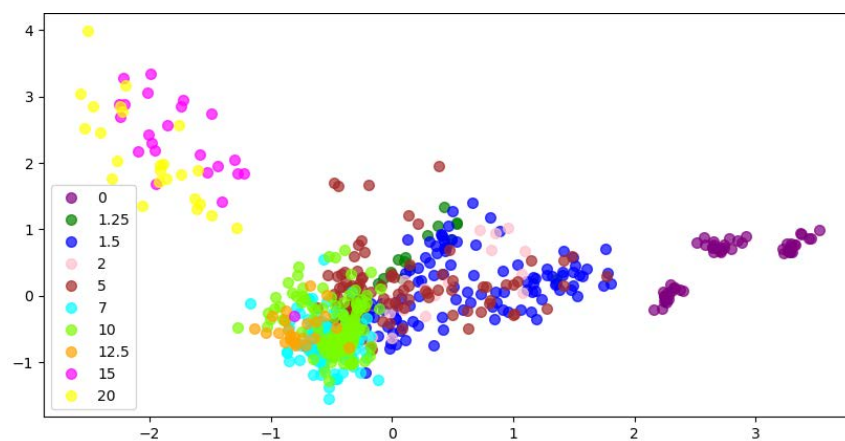


Figura G.14: Representação através de PCA das concentrações de glucose (mM), baseada nos componentes L e S do espaço HLS.



## G.2 Quatro Níveis de Glucose

### G.2.1 Estado de jejum

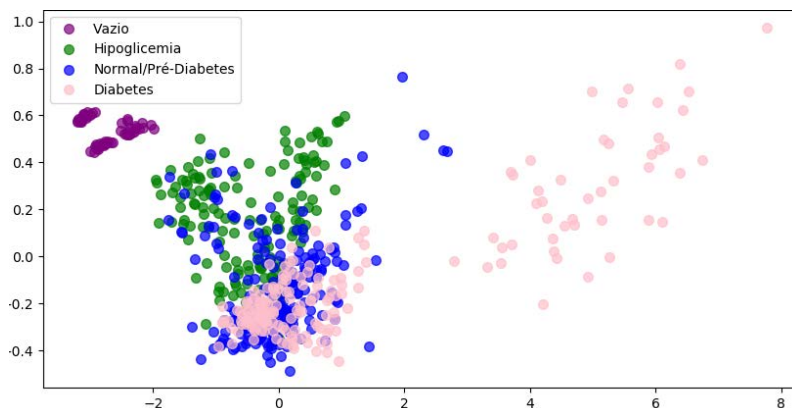


Figura G.15: Representação através de PCA dos quatro níveis de referência da glucose (em jejum), baseada na média dos componentes RGB.

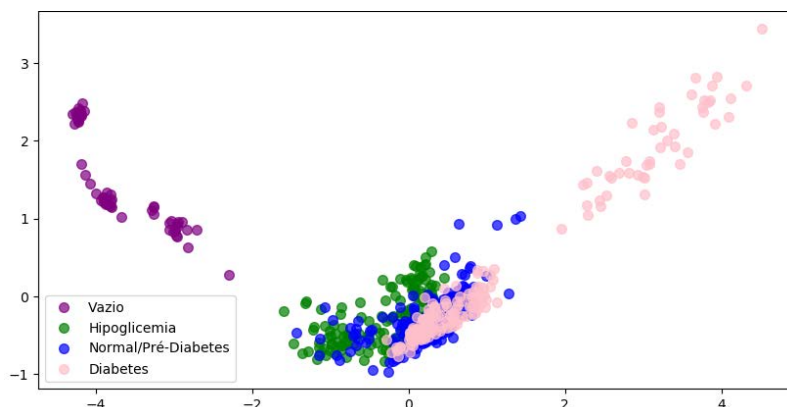


Figura G.16: Representação através de PCA dos quatro níveis de referência da glucose (em jejum), baseada na média dos componentes HSV.

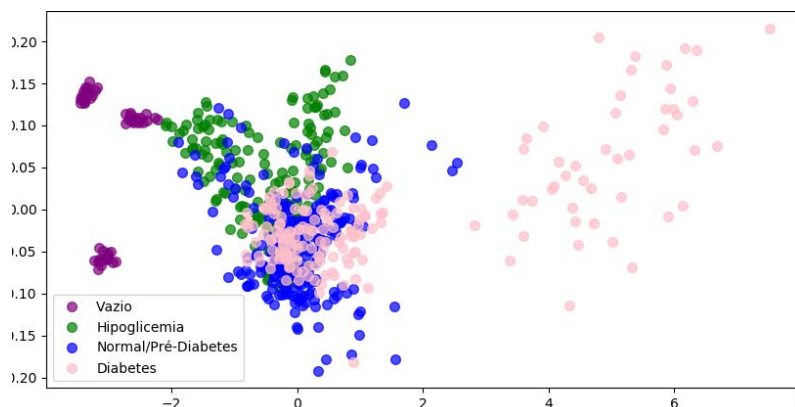


Figura G.17: Representação através de PCA dos quatro níveis de referência da glucose (em jejum), baseada na média dos componentes XYZ.

APÊNDICE G. RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

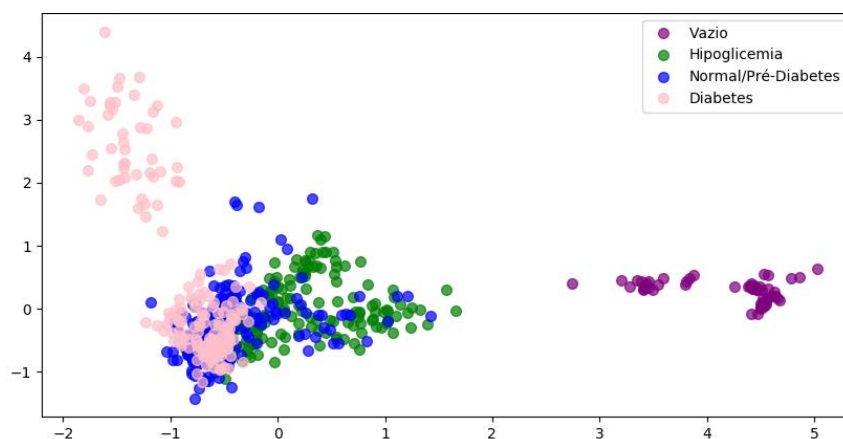


Figura G.18: Representação através de PCA dos quatro níveis de referência da glucose (em jejum), baseada na média dos componentes HLS.

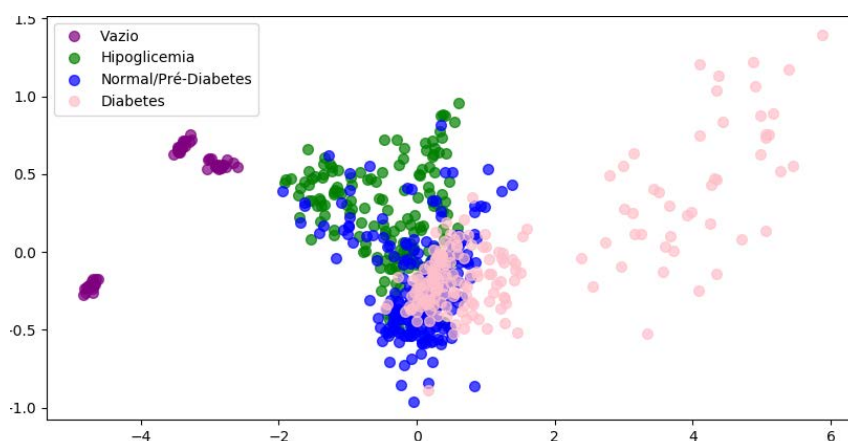


Figura G.19: Representação através de PCA dos quatro níveis de referência da glucose (em jejum), baseada na média dos componentes LAB.

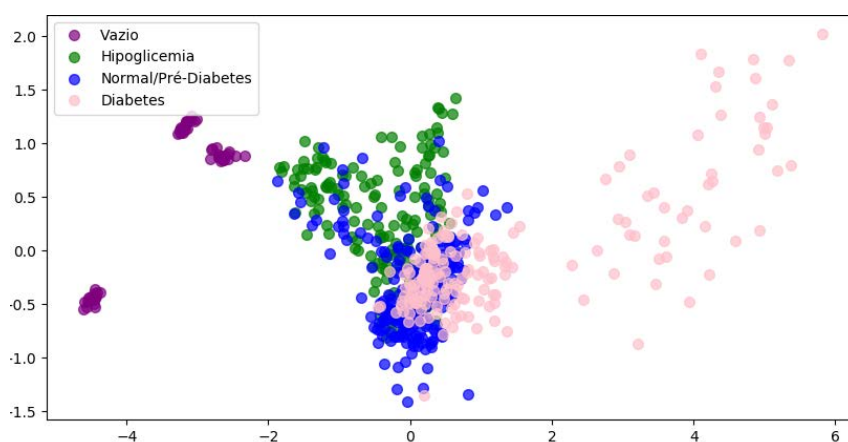


Figura G.20: Representação através de PCA dos quatro níveis de referência da glucose (em jejum), baseada na média dos componentes YUV.

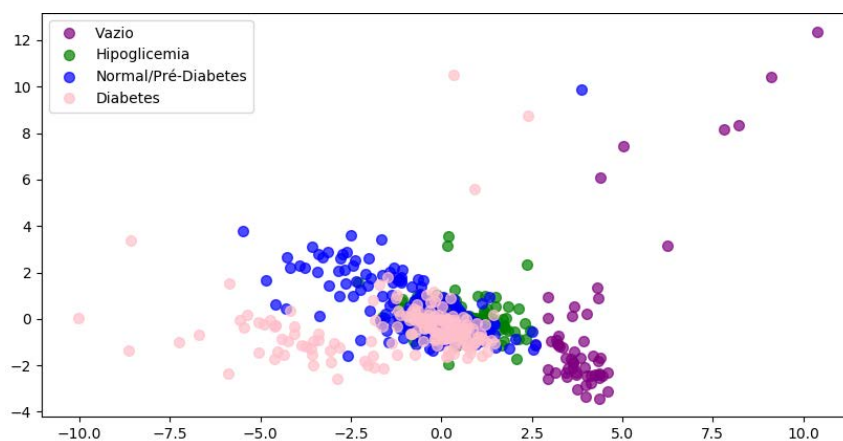


Figura G.21: Representação através de PCA dos quatro níveis de referência da glucose (em jejum), baseada na média, desvio padrão e assimetria dos componentes RGB.

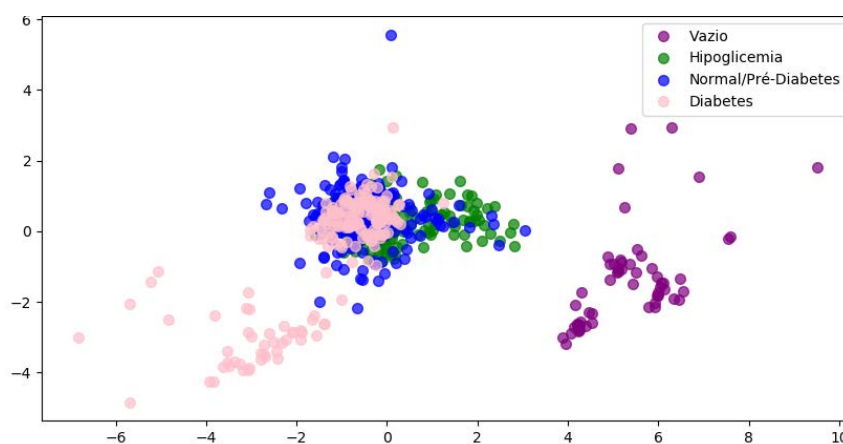


Figura G.22: Representação através de PCA dos quatro níveis de referência da glucose (em jejum), baseada na média, desvio padrão e assimetria dos componentes HSV.

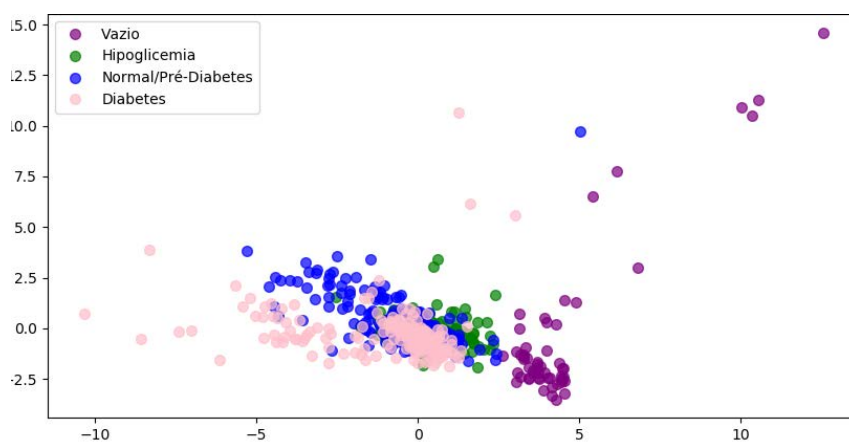


Figura G.23: Representação através de PCA dos quatro níveis de referência da glucose (em jejum), baseada na média, desvio padrão e assimetria dos componentes XYZ.

APÊNDICE G. RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

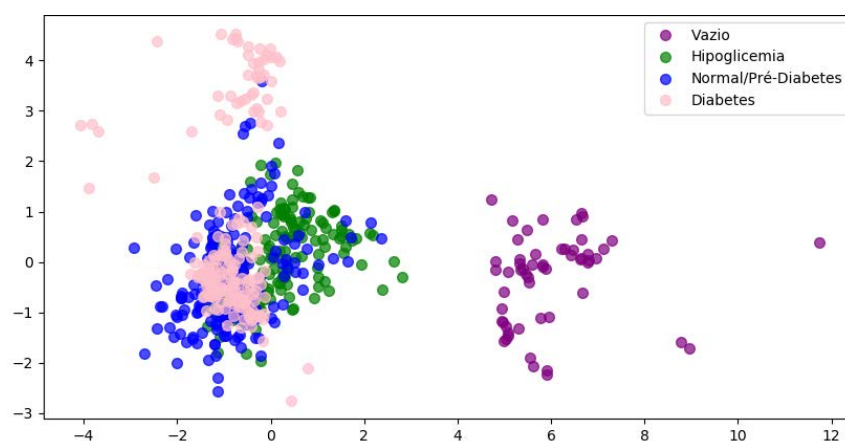


Figura G.24: Representação através de PCA dos quatro níveis de referência da glucose (em jejum), baseada na média, desvio padrão e assimetria dos componentes HLS.

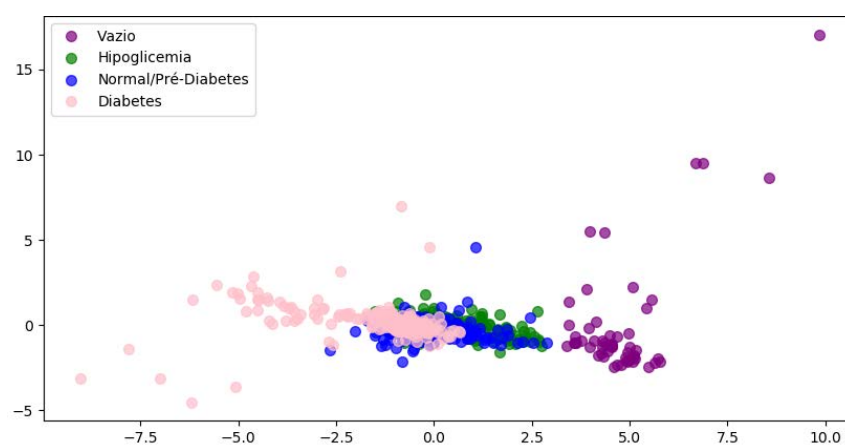


Figura G.25: Representação através de PCA dos quatro níveis de referência da glucose (em jejum), baseada na média, desvio padrão e assimetria dos componentes LAB.

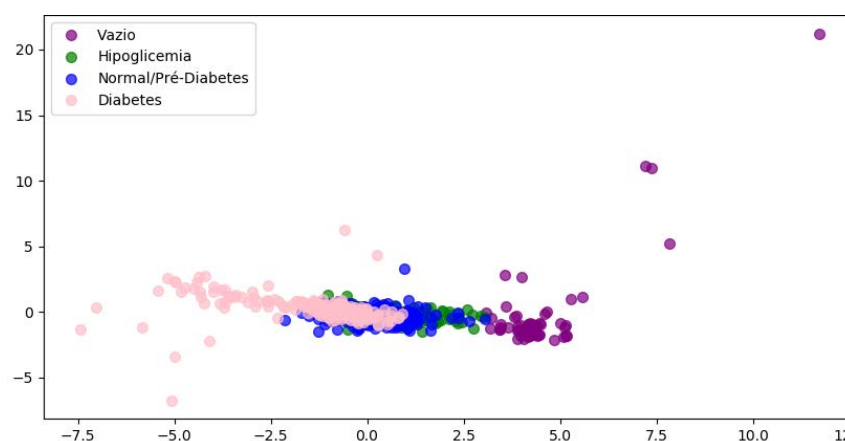


Figura G.26: Representação através de PCA dos quatro níveis de referência da glucose (em jejum), baseada na média, desvio padrão e assimetria dos componentes YUV.

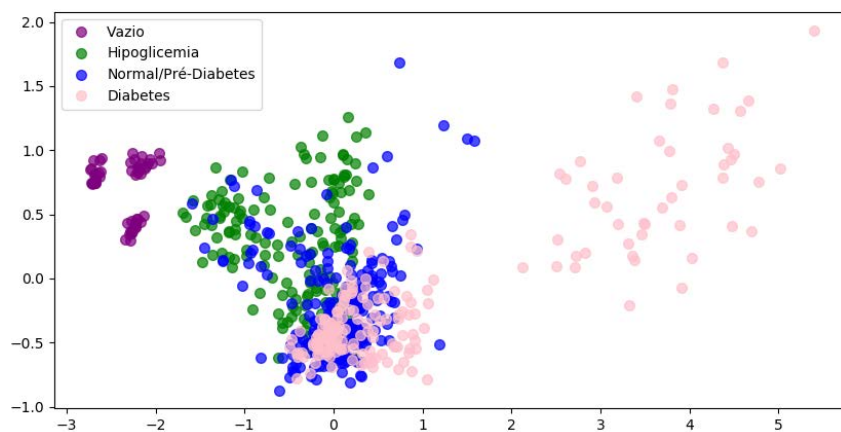


Figura G.27: Representação através de PCA dos quatro níveis de referência da glucose (em jejum), baseada nos componentes S e V do espaço HSV.

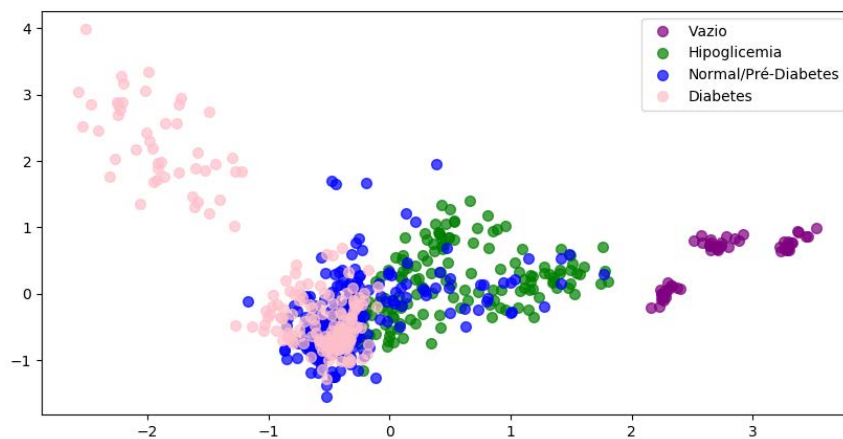


Figura G.28: Representação através de PCA dos quatro níveis de referência da glucose (em jejum), baseada nos componentes L e S do espaço HLS.

## G.2.2 Estado de duas horas após a refeição

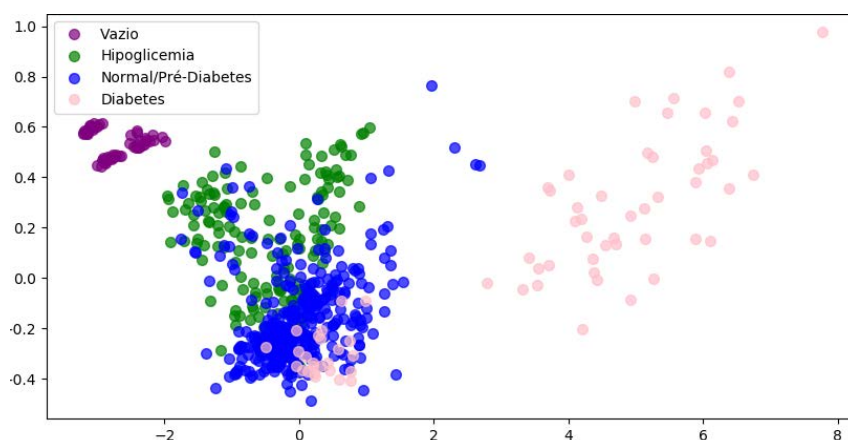


Figura G.29: Representação através de PCA dos quatro níveis de referência da glucose (duas horas após a refeição), baseada na média dos componentes RGB.

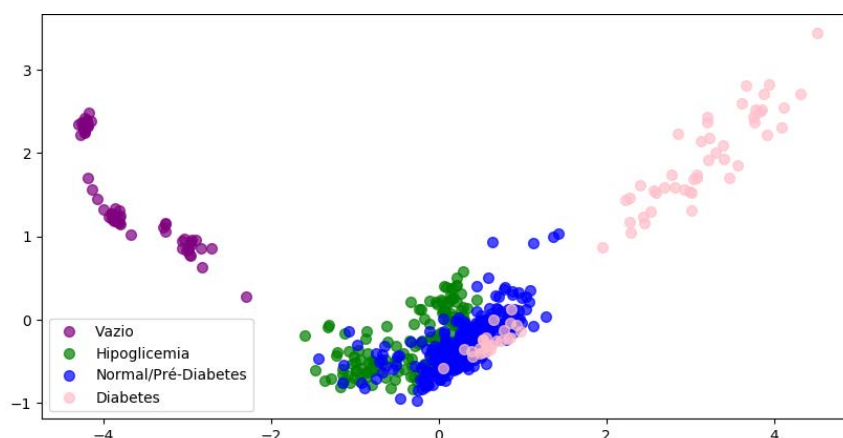


Figura G.30: Regiões de decisão para os quatro níveis de referência da glucose (duas horas após a refeição) dos três algoritmos estudados, baseados na média dos componentes HSV.

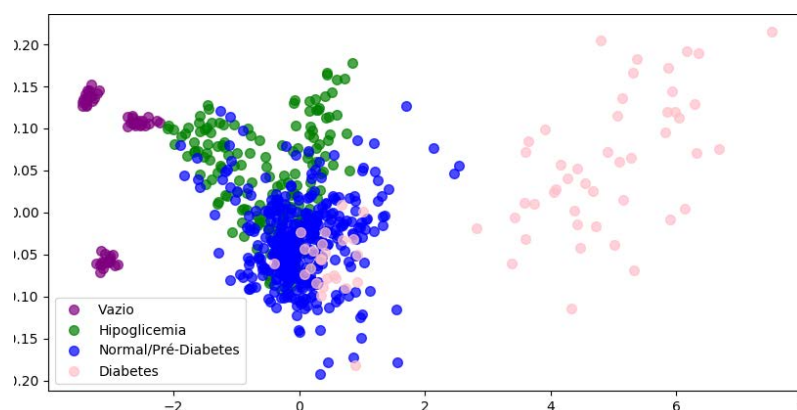


Figura G.31: Representação através de PCA dos quatro níveis de referência da glucose (duas horas após a refeição), baseada na média dos componentes XYZ.

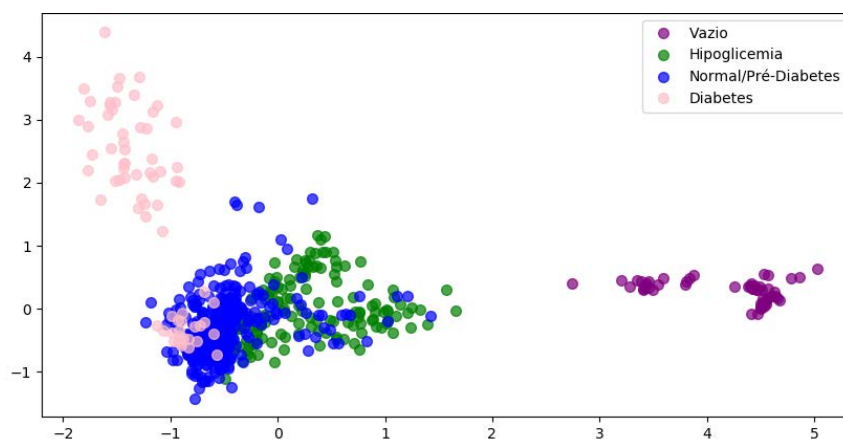


Figura G.32: Representação através de PCA dos quatro níveis de referência da glucose (duas horas após a refeição), baseada na média dos componentes HLS.

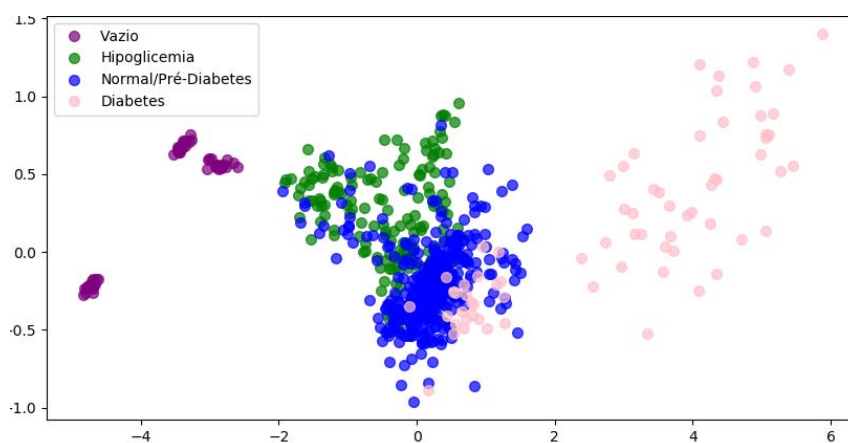


Figura G.33: Representação através de PCA dos quatro níveis de referência da glucose (duas horas após a refeição), baseada na média dos componentes LAB.

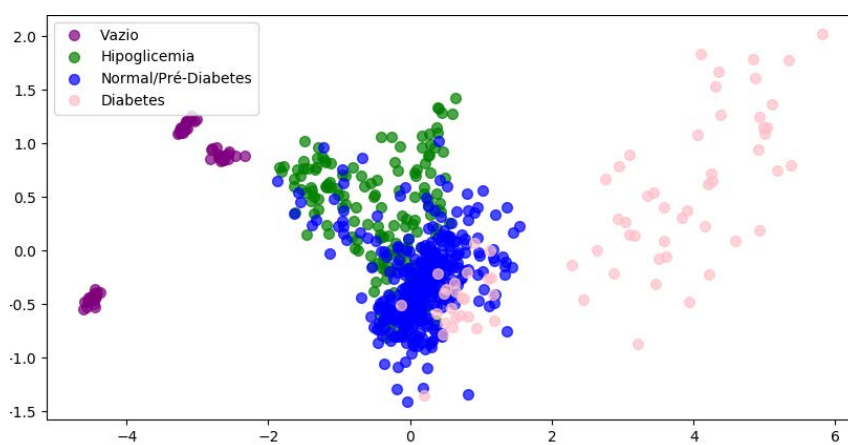


Figura G.34: Representação através de PCA dos quatro níveis de referência da glucose (duas horas após a refeição), baseada na média dos componentes YUV.



APÊNDICE G. RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

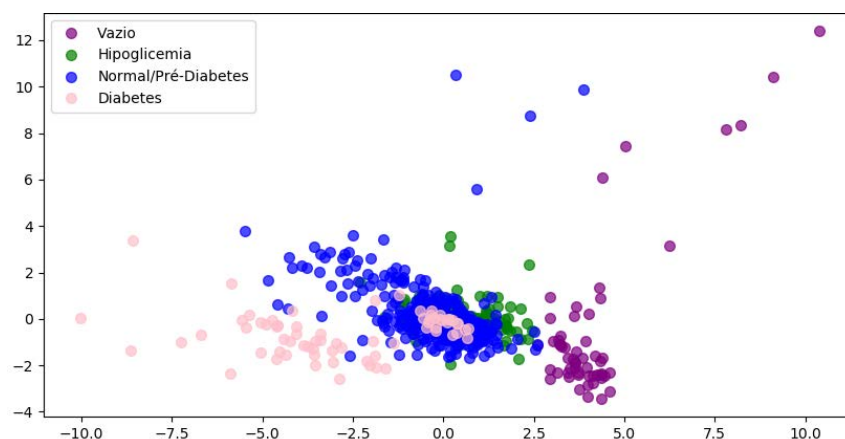


Figura G.35: Representação através de PCA dos quatro níveis de referência da glucose (duas horas após a refeição), baseada na média, desvio padrão e assimetria dos componentes RGB.

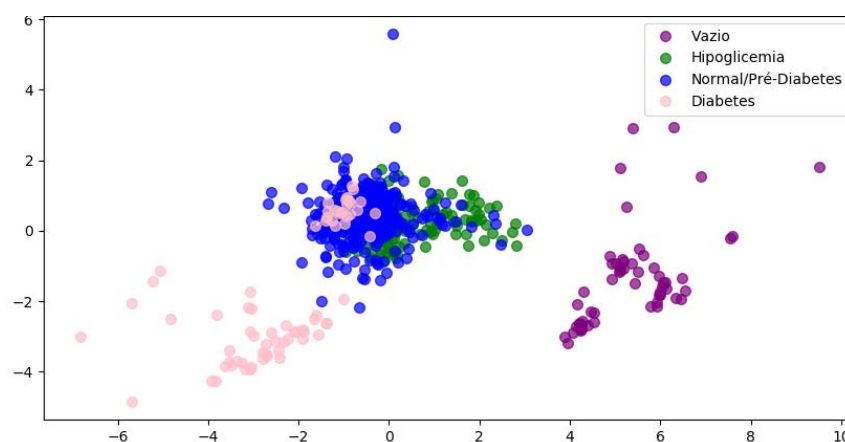


Figura G.36: Representação através de PCA dos quatro níveis de referência da glucose (duas horas após a refeição), baseada na média, desvio padrão e assimetria dos componentes HSV.

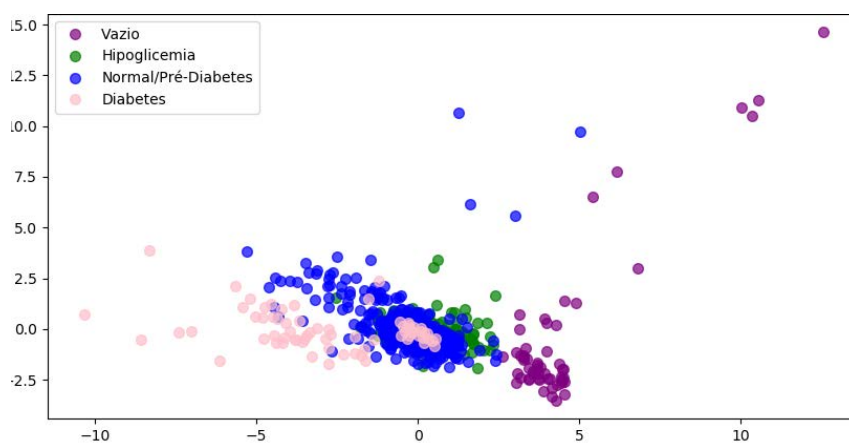


Figura G.37: Representação através de PCA dos quatro níveis de referência da glucose (duas horas após a refeição), baseada na média, desvio padrão e assimetria dos componentes XYZ.



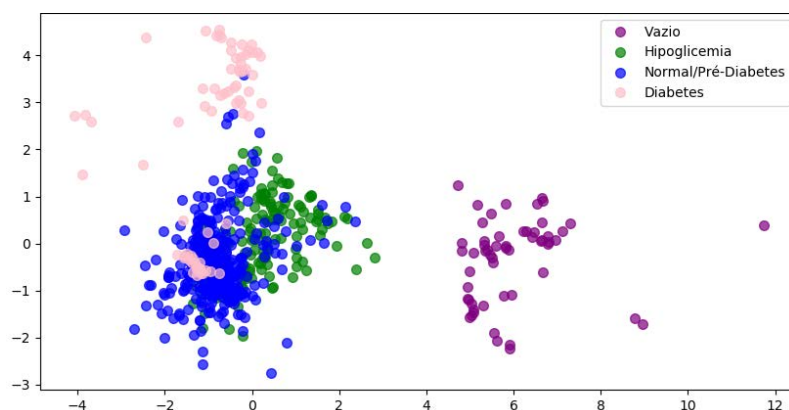


Figura G.38: Representação através de PCA dos quatro níveis de referência da glucose (duas horas após a refeição), baseada na média, desvio padrão e assimetria dos componentes HLS.

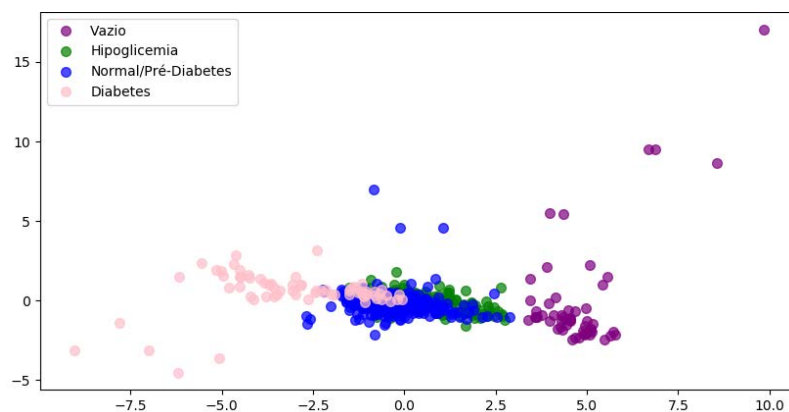


Figura G.39: Representação através de PCA dos quatro níveis de referência da glucose (duas horas após a refeição), baseada na média, desvio padrão e assimetria dos componentes LAB.

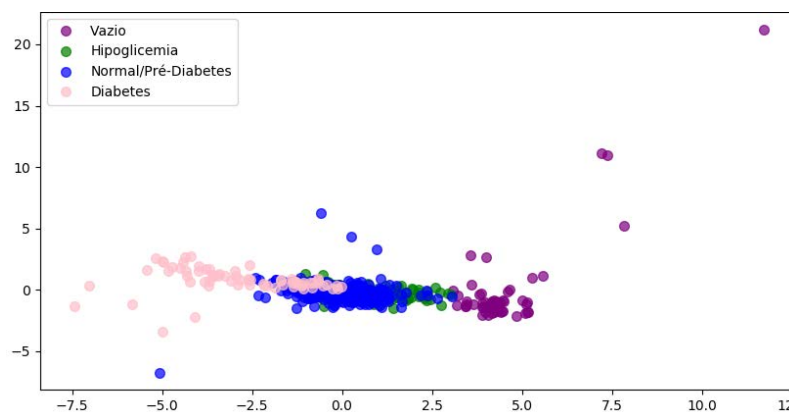


Figura G.40: Representação através de PCA dos quatro níveis de referência da glucose (duas horas após a refeição), baseada na média, desvio padrão e assimetria dos componentes YUV.

APÊNDICE G. RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

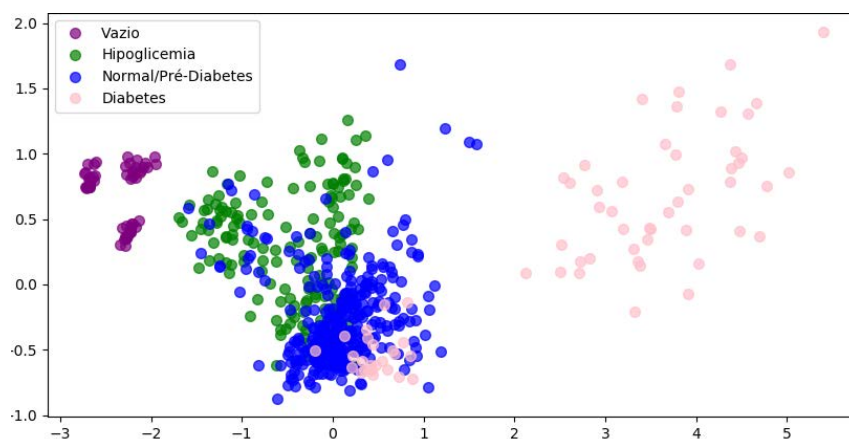


Figura G.41: Representação através de PCA dos quatro níveis de referência da glucose (duas horas após a refeição), baseada nos componentes S e V do espaço HSV.

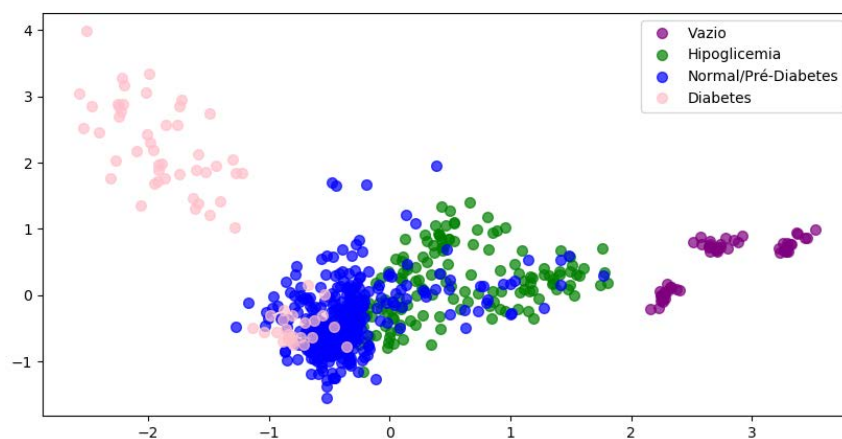


Figura G.42: Representação através de PCA dos quatro níveis de referência da glucose (duas horas após a refeição), baseada nos componentes L e S do espaço HLS.

## G.3 Cinco Níveis de Glucose

### G.3.1 Estado de jejum

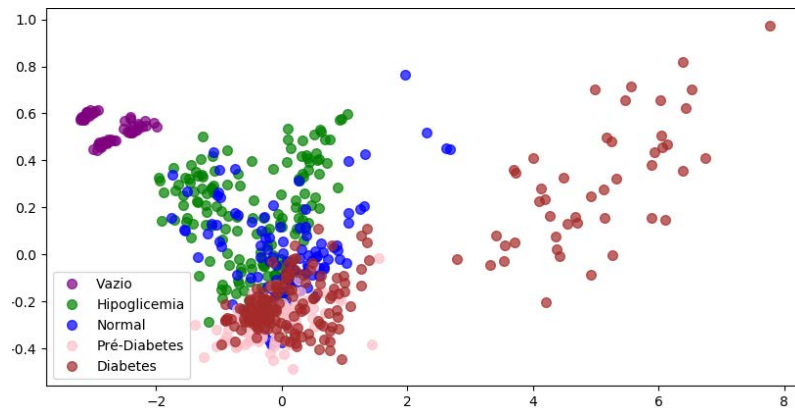


Figura G.43: Representação através de PCA dos cinco níveis de referência da glucose (em jejum), baseada na média dos componentes RGB.

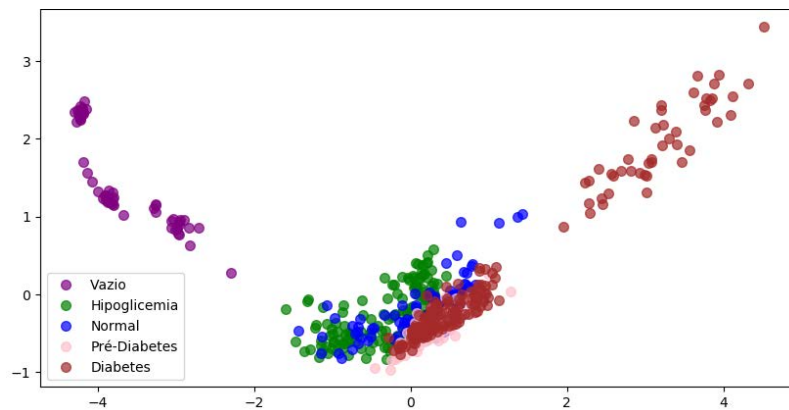


Figura G.44: Representação através de PCA dos cinco níveis de referência da glucose (em jejum), baseada na média dos componentes HSV.

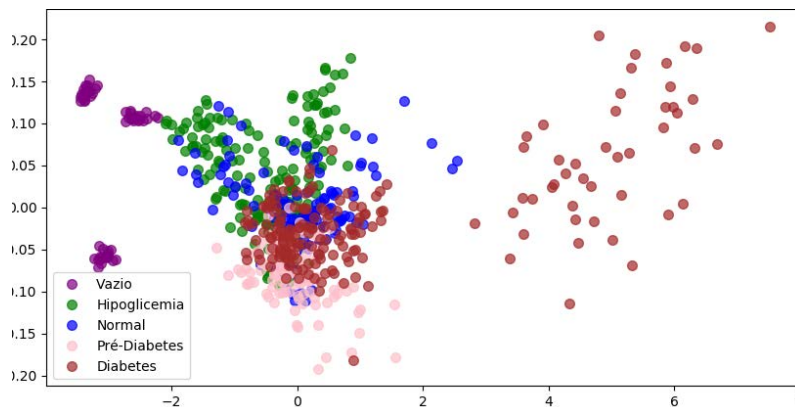


Figura G.45: Representação através de PCA dos cinco níveis de referência da glucose (em jejum), baseada na média dos componentes XYZ.

APÊNDICE G. RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

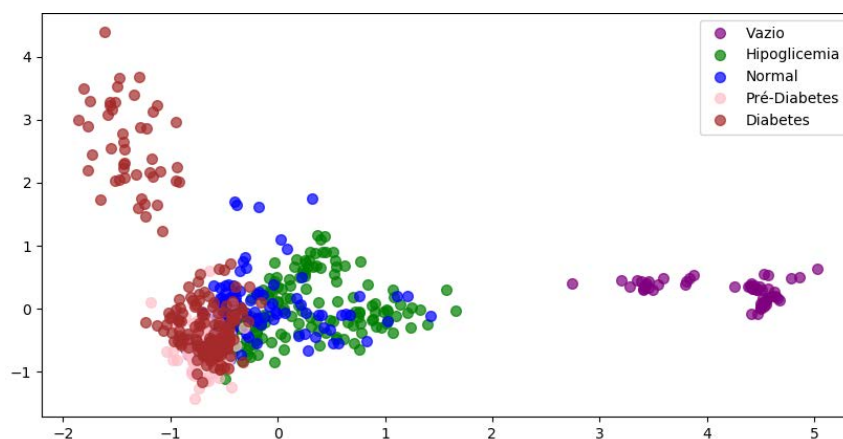


Figura G.46: Representação através de PCA dos cinco níveis de referência da glucose (em jejum), baseada na média dos componentes HLS.

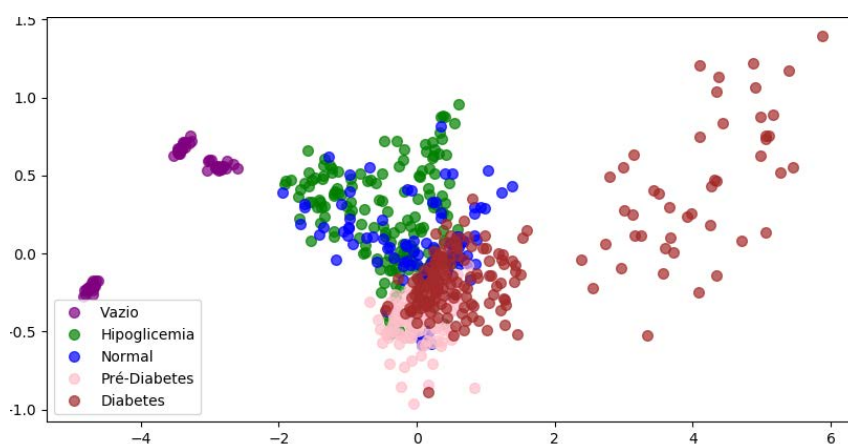


Figura G.47: Representação através de PCA dos cinco níveis de referência da glucose (em jejum), baseada na média dos componentes LAB.

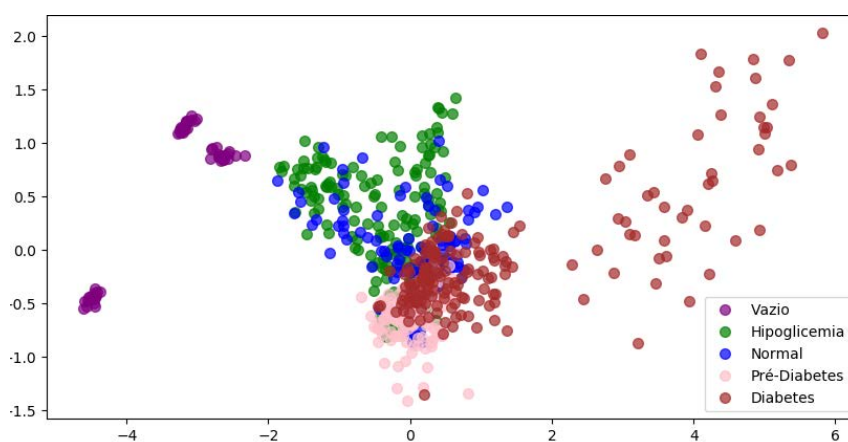


Figura G.48: Representação através de PCA dos cinco níveis de referência da glucose (em jejum), baseada na média dos componentes YUV.

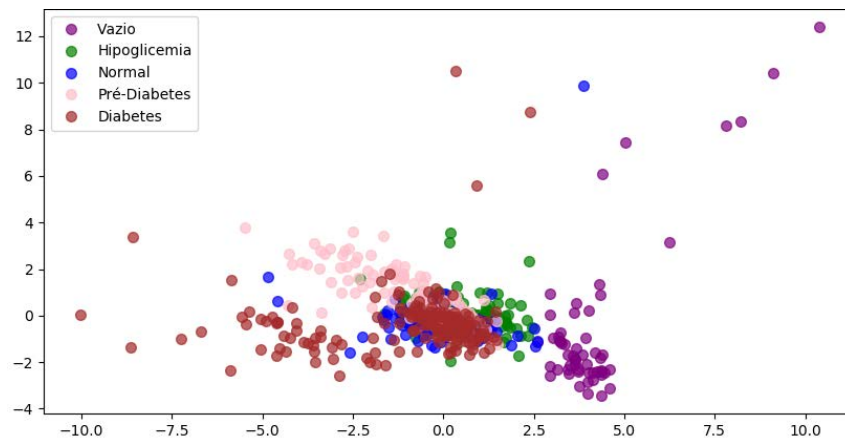


Figura G.49: Representação através de PCA dos cinco níveis de referência da glucose (em jejum), baseada na média, desvio padrão e assimetria dos componentes RGB.

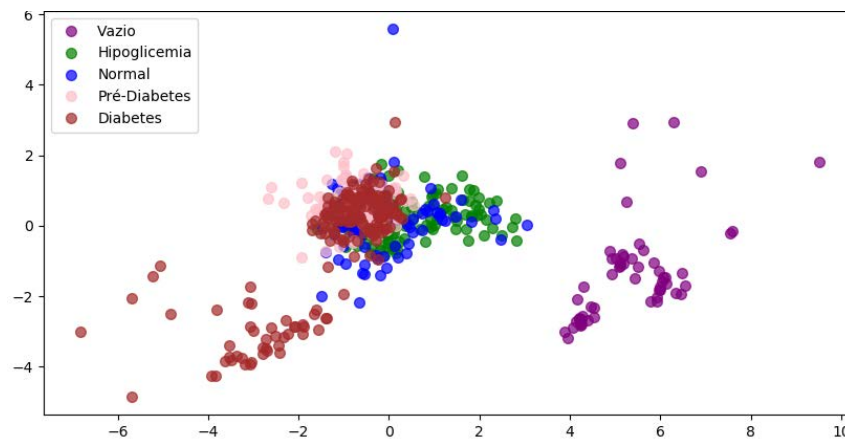


Figura G.50: Representação através de PCA dos cinco níveis de referência da glucose (em jejum), baseada na média, desvio padrão e assimetria dos componentes HSV.

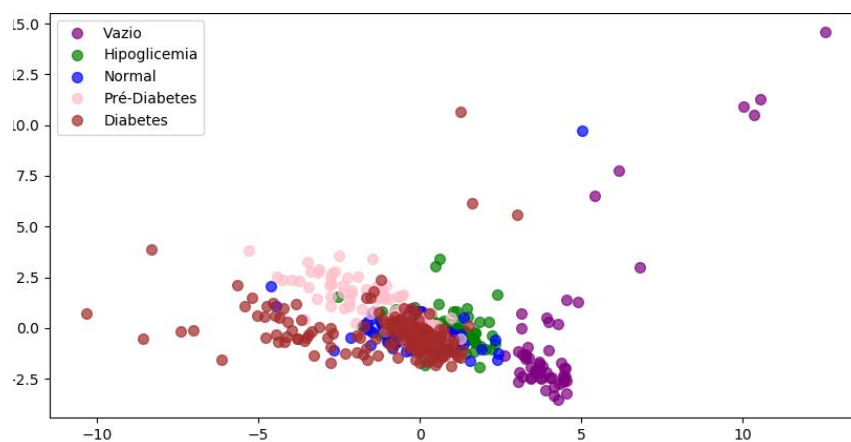


Figura G.51: Representação através de PCA dos cinco níveis de referência da glucose (em jejum), baseada na média, desvio padrão e assimetria dos componentes XYZ.

APÊNDICE G. RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

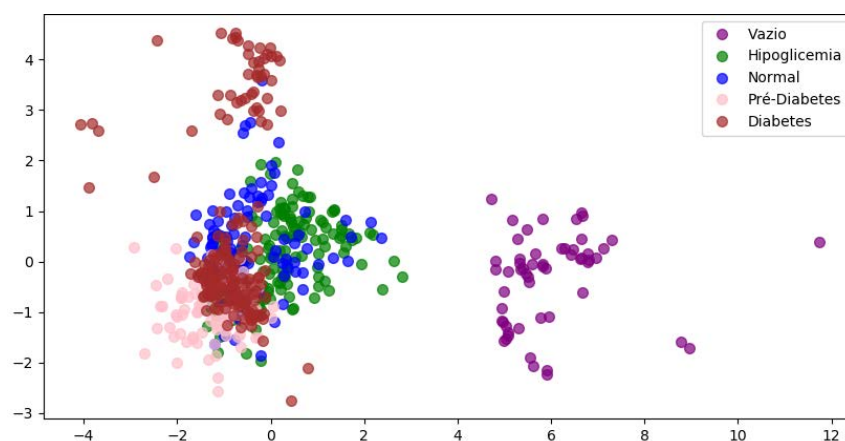


Figura G.52: Representação através de PCA dos cinco níveis de referência da glucose (em jejum), baseada na média, desvio padrão e assimetria dos componentes HLS.

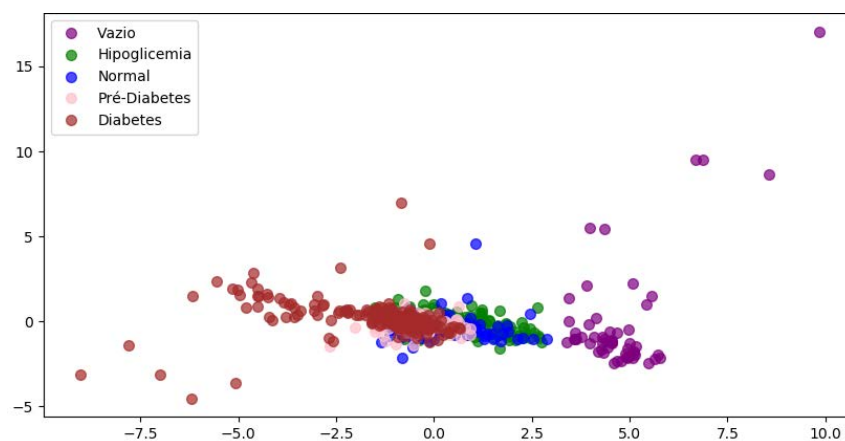


Figura G.53: Representação através de PCA dos cinco níveis de referência da glucose (em jejum), baseada na média, desvio padrão e assimetria dos componentes LAB.

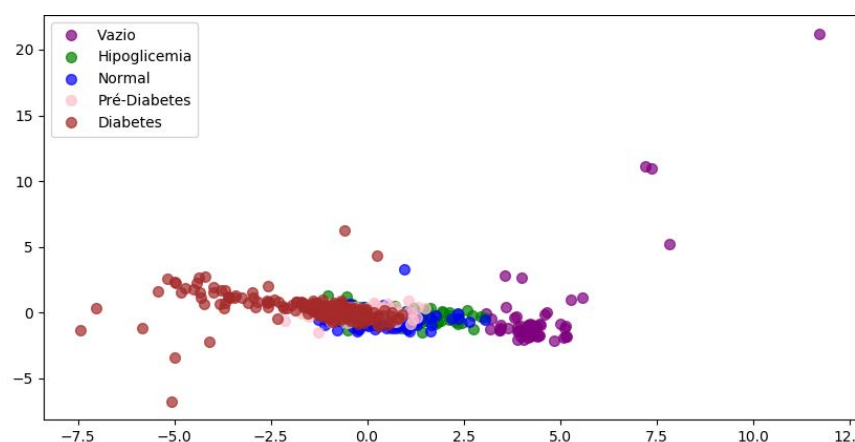


Figura G.54: Representação através de PCA dos cinco níveis de referência da glucose (em jejum), baseada na média, desvio padrão e assimetria dos componentes YUV.

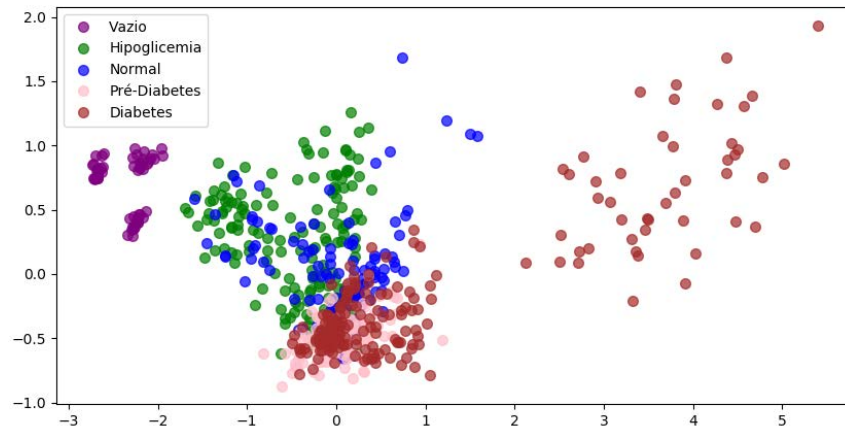


Figura G.55: Representação através de PCA dos cinco níveis de referência da glucose (em jejum), baseada nos componentes S e V do espaço HSV.

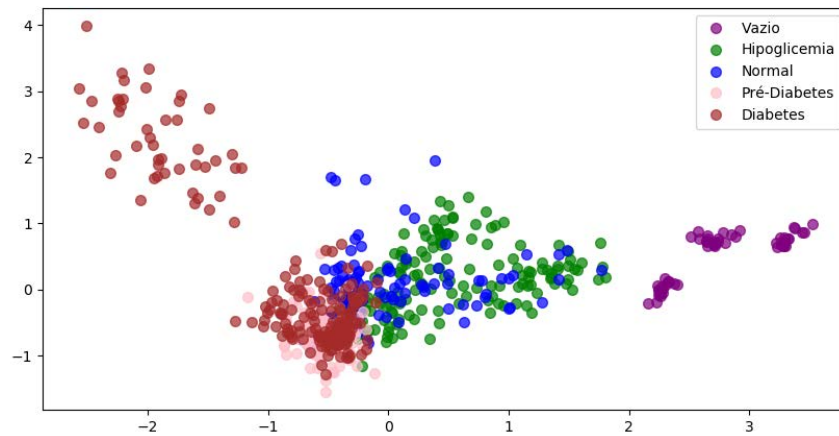


Figura G.56: Representação através de PCA dos cinco níveis de referência da glucose (em jejum), baseada nos componentes L e S do espaço HLS.



### G.3.2 Estado de duas horas após a refeição

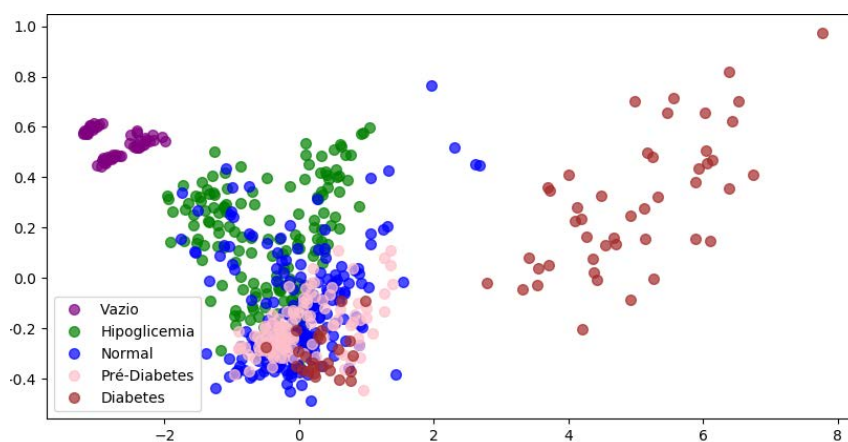


Figura G.57: Representação através de PCA dos cinco níveis de referência da glucose (duas horas após a refeição), baseada na média dos componentes RGB.

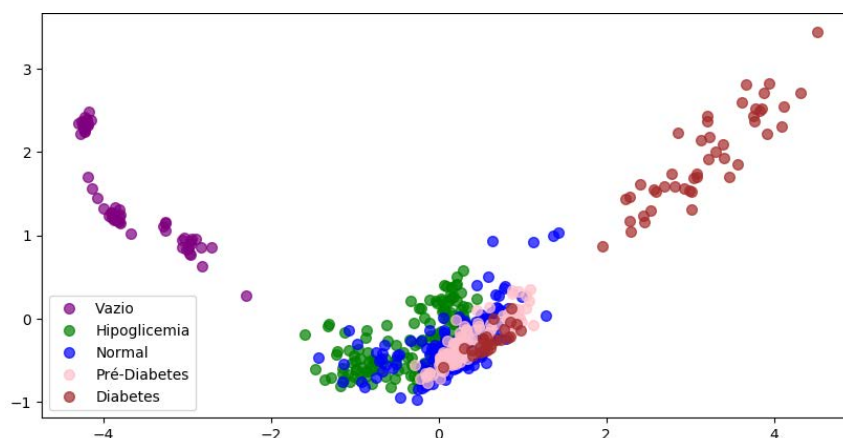


Figura G.58: Representação através de PCA dos cinco níveis de referência da glucose (duas horas após a refeição), baseada na média dos componentes HSV.

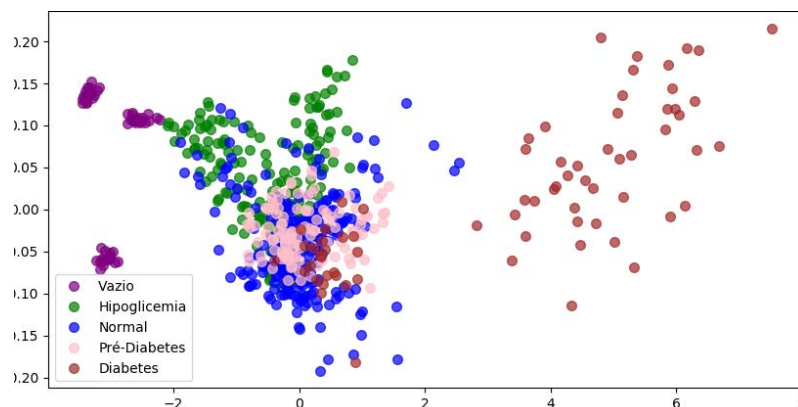


Figura G.59: Representação através de PCA dos cinco níveis de referência da glucose (duas horas após a refeição), baseada na média dos componentes XYZ.



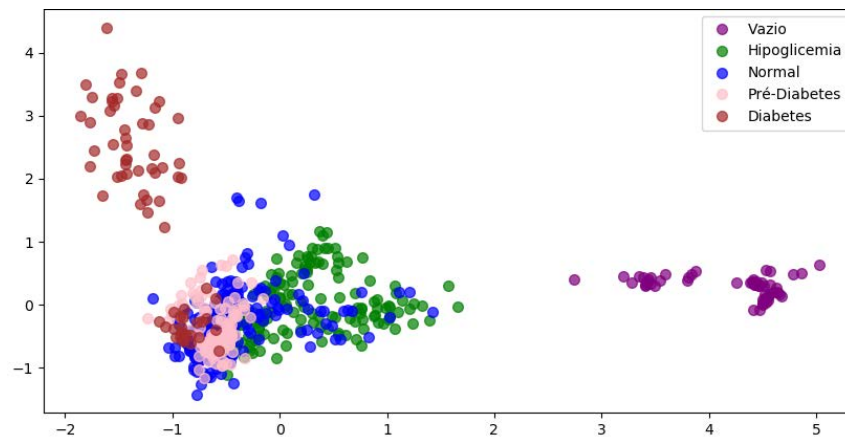


Figura G.60: Representação através de PCA dos cinco níveis de referência da glucose (duas horas após a refeição), baseada na média dos componentes HLS.

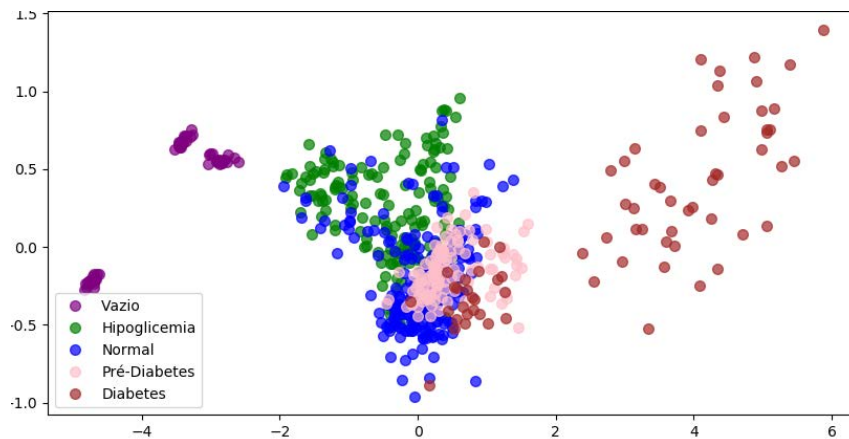


Figura G.61: Representação através de PCA dos cinco níveis de referência da glucose (duas horas após a refeição), baseada na média dos componentes LAB.

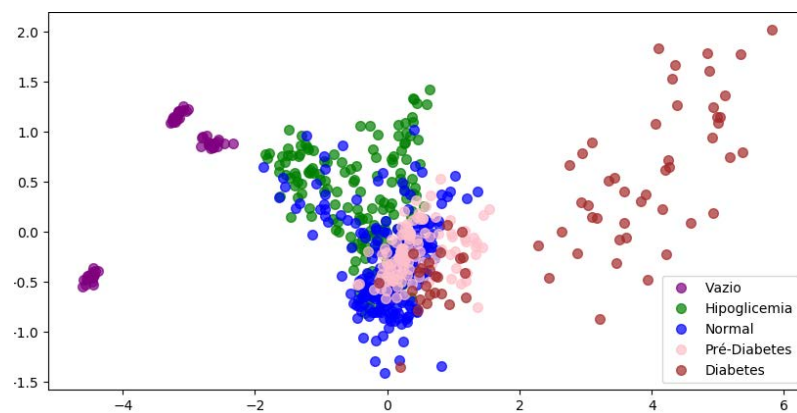


Figura G.62: Representação através de PCA dos cinco níveis de referência da glucose (duas horas após a refeição), baseada na média dos componentes YUV.

## APÊNDICE G. RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

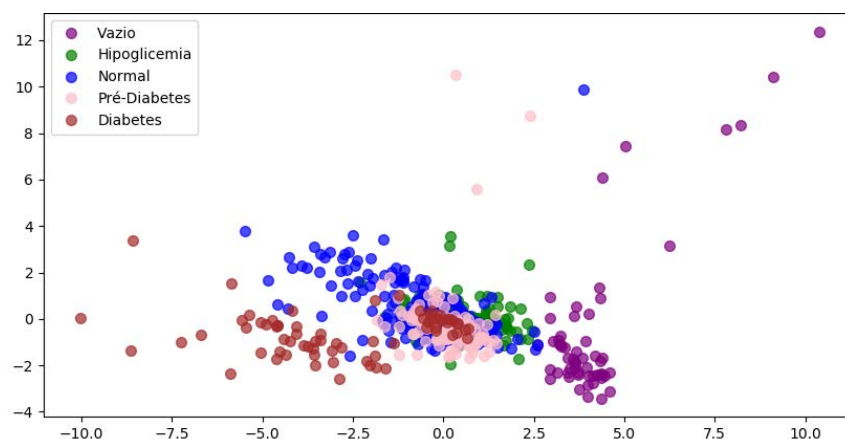


Figura G.63: Representação através de PCA dos cinco níveis de referência da glucose (duas horas após a refeição), baseada na média, desvio padrão e assimetria dos componentes RGB.

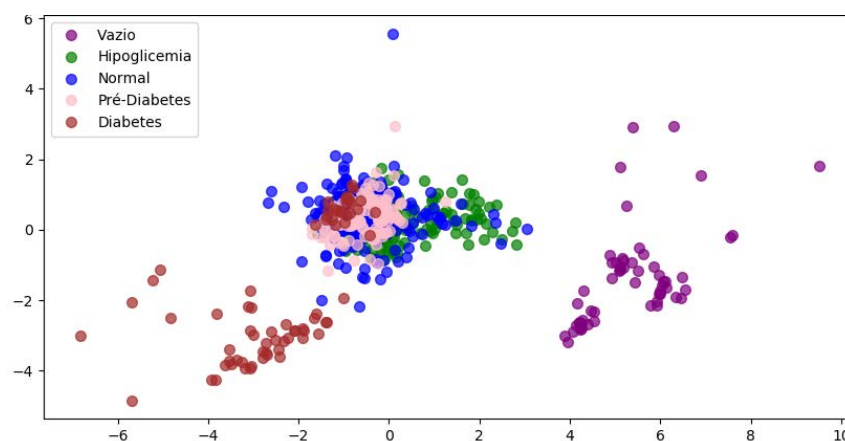


Figura G.64: Representação através de PCA dos cinco níveis de referência da glucose (duas horas após a refeição), baseada na média, desvio padrão e assimetria dos componentes HSV.

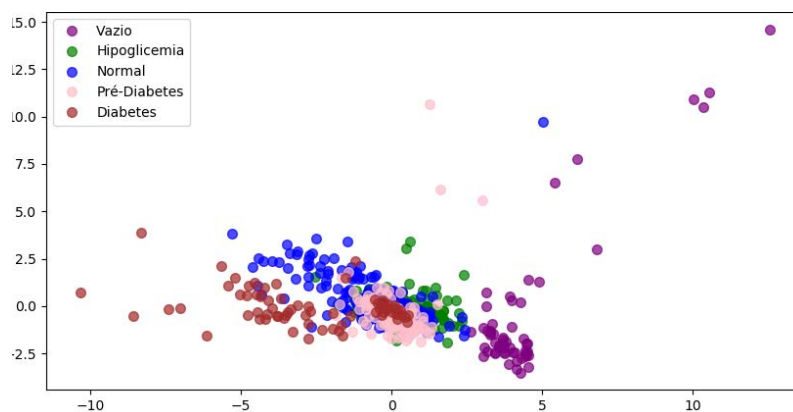


Figura G.65: Representação através de PCA dos cinco níveis de referência da glucose (duas horas após a refeição), baseada na média, desvio padrão e assimetria dos componentes XYZ.

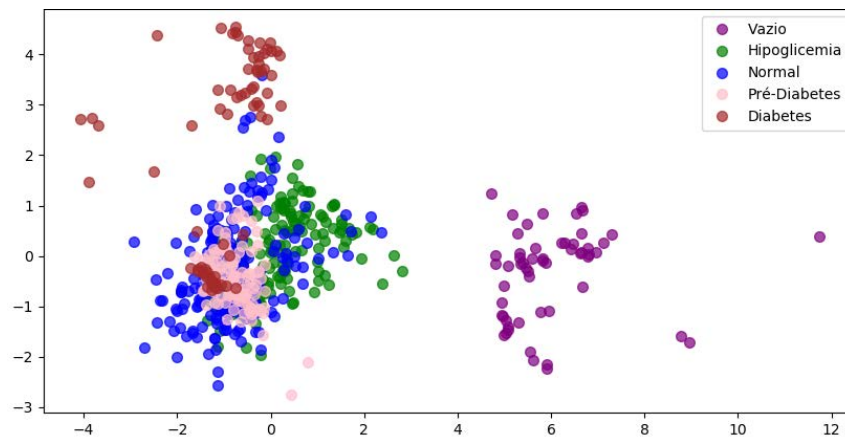


Figura G.66: Representação através de PCA dos cinco níveis de referência da glucose (duas horas após a refeição), baseada na média, desvio padrão e assimetria dos componentes HLS.

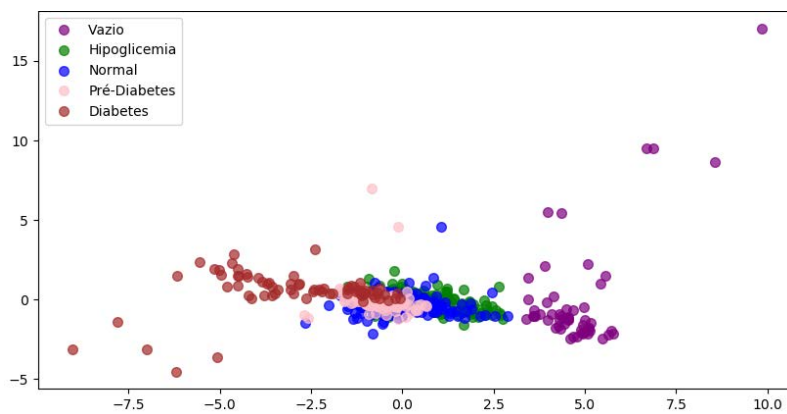


Figura G.67: Representação através de PCA dos cinco níveis de referência da glucose (duas horas após a refeição), baseada na média, desvio padrão e assimetria dos componentes LAB.

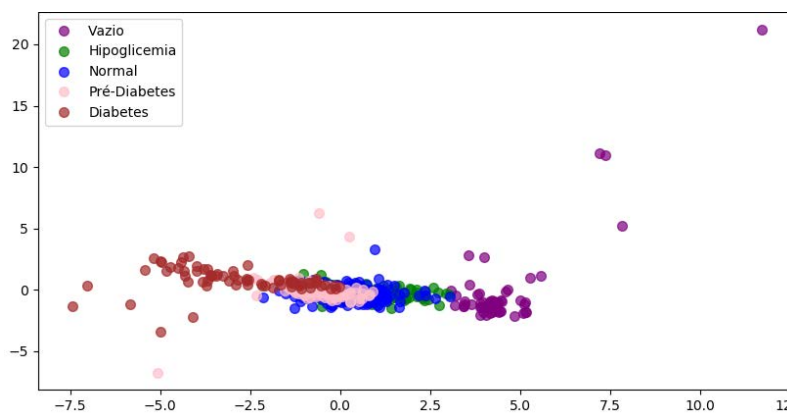


Figura G.68: Representação através de PCA dos cinco níveis de referência da glucose (duas horas após a refeição), baseada na média, desvio padrão e assimetria dos componentes YUV.

APÊNDICE G. RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

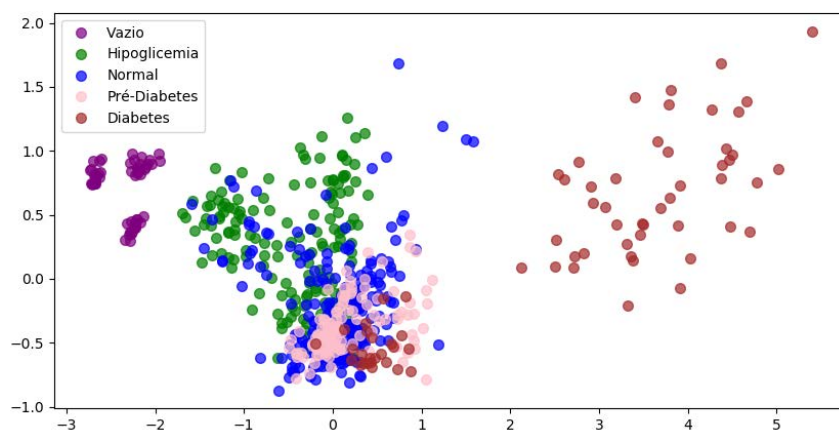


Figura G.69: Representação através de PCA dos cinco níveis de referência da glucose (duas horas após a refeição), baseada nos componentes S e V do espaço HSV.

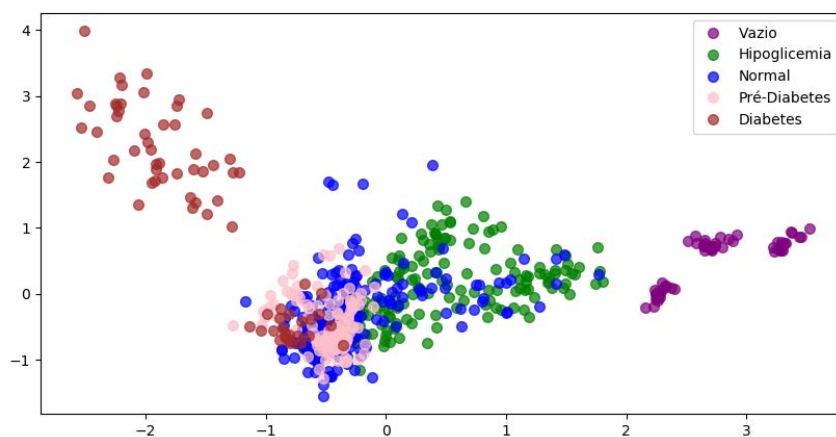


Figura G.70: Representação através de PCA dos cinco níveis de referência da glucose (duas horas após a refeição), baseada nos componentes L e S do espaço HLS.

## G.4 Concentrações de Ácido Úrico

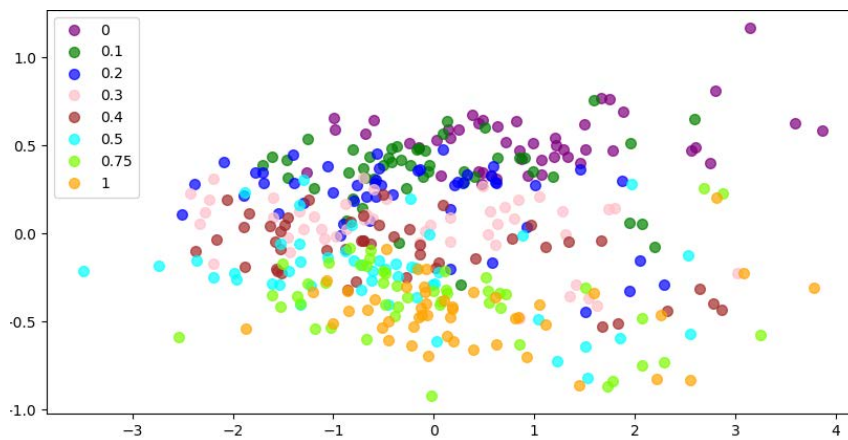


Figura G.71: Representação através de PCA das concentrações do ácido úrico (mM), baseada na média dos componentes RGB.

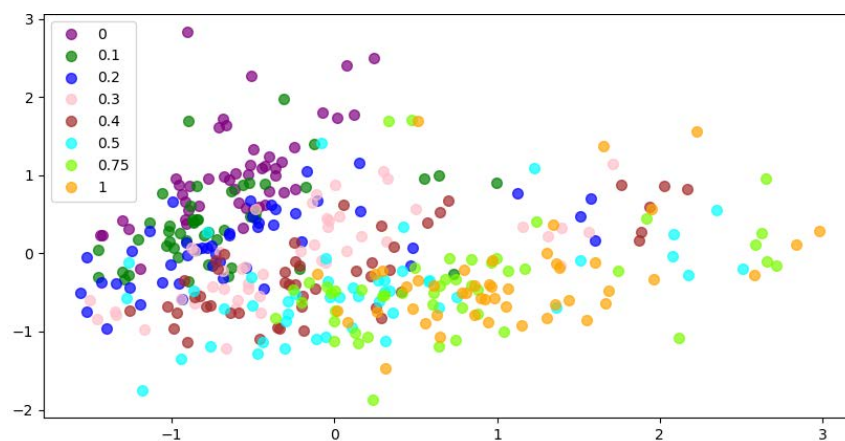


Figura G.72: Representação através de PCA das concentrações do ácido úrico (mM), baseada na média dos componentes HSV.

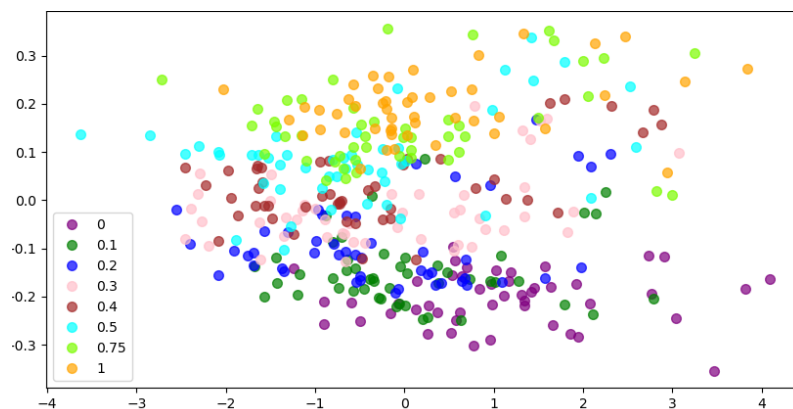


Figura G.73: Representação através de PCA das concentrações do ácido úrico (mM), baseada na média dos componentes XYZ.

APÊNDICE G. RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

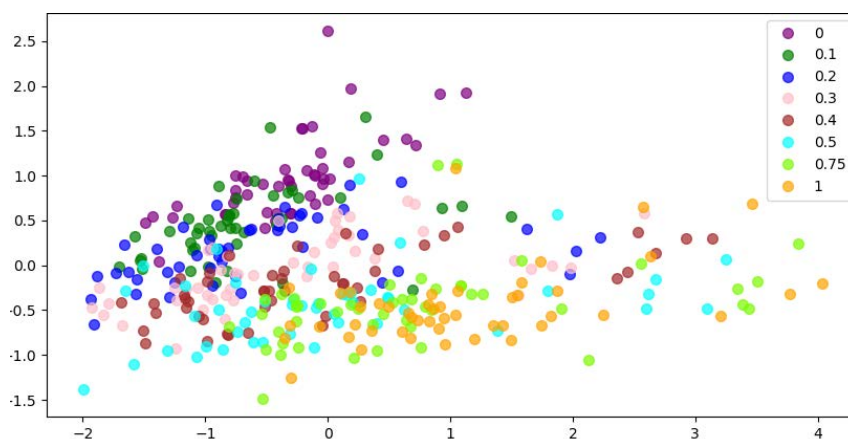


Figura G.74: Representação através de PCA das concentrações do ácido úrico (mM), baseada na média dos componentes HLS.

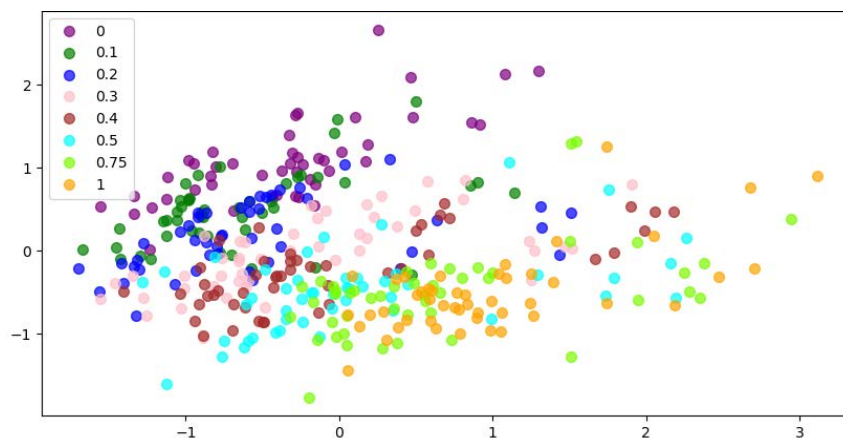


Figura G.75: Representação através de PCA das concentrações do ácido úrico (mM), baseada na média dos componentes LAB.

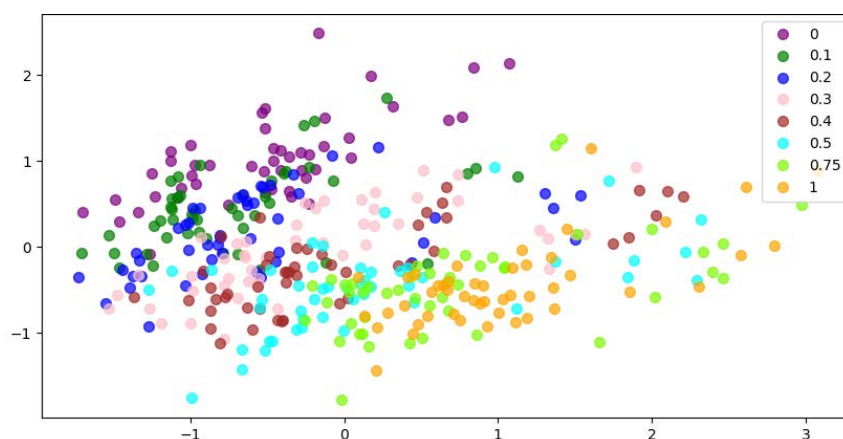


Figura G.76: Representação através de PCA das concentrações do ácido úrico (mM), baseada na média dos componentes YUV.



#### G.4. CONCENTRAÇÕES DE ÁCIDO ÚRICO

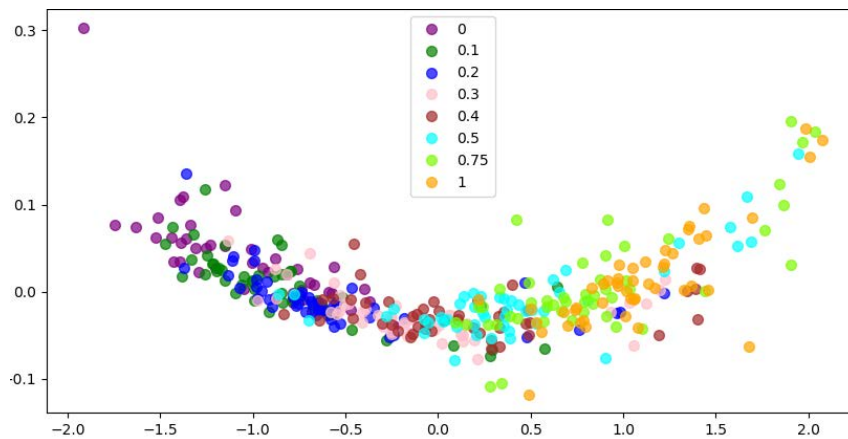


Figura G.77: Representação através de PCA das concentrações do ácido úrico (mM), baseada no rácio entre os componentes vermelho e azul do espaço RGB e componente H do espaço HSV.

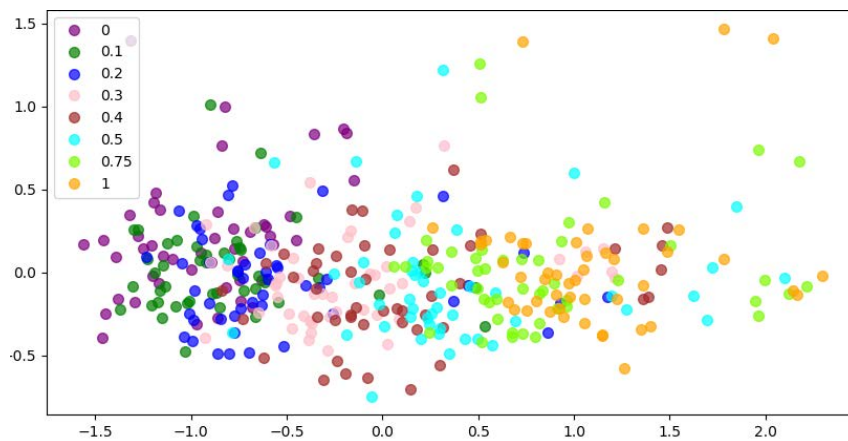


Figura G.78: Representação através de PCA das concentrações do ácido úrico (mM), baseada nos componentes a e b do espaço LAB.

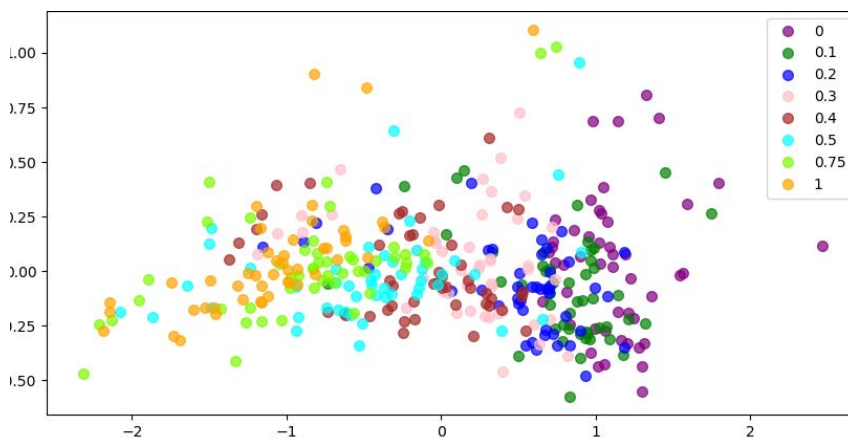


Figura G.79: Representação através de PCA das concentrações do ácido úrico (mM), baseada nos componentes H e S do espaço HSV.

APÊNDICE G. RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

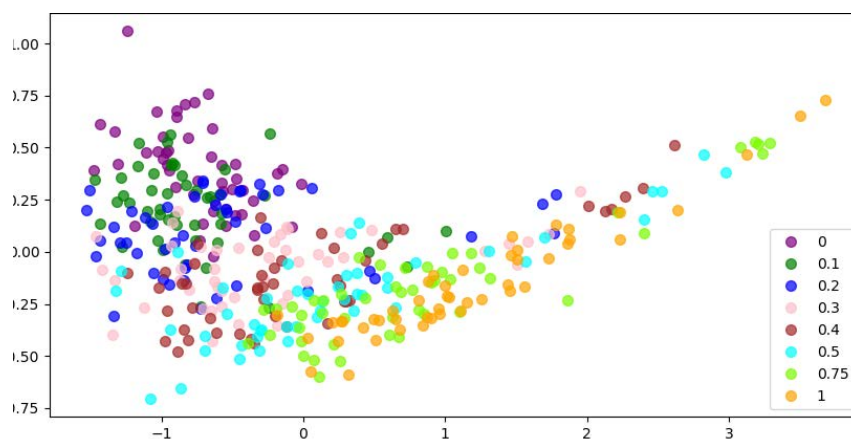


Figura G.80: Representação através de PCA das concentrações do ácido úrico (mM), baseada nos componentes H e S do espaço HLS.

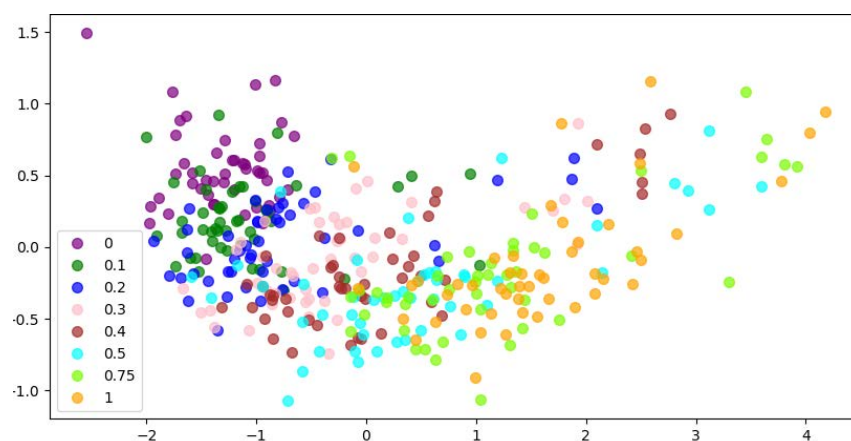


Figura G.81: Representação através de PCA das concentrações do ácido úrico (mM), baseada nos componentes H e S dos espaços HSV e HLS.

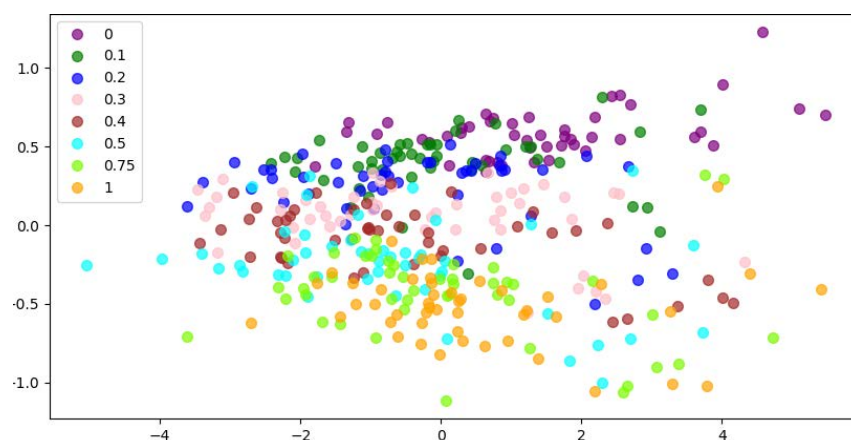


Figura G.82: Representação através de PCA das concentrações do ácido úrico (mM), baseada na combinação da média dos componentes do modelo RGB com as componentes V do modelo HSV, Z do modelo XYZ, A do modelo LAB e Y do modelo YUV.



## G.5 Níveis de Ácido Úrico

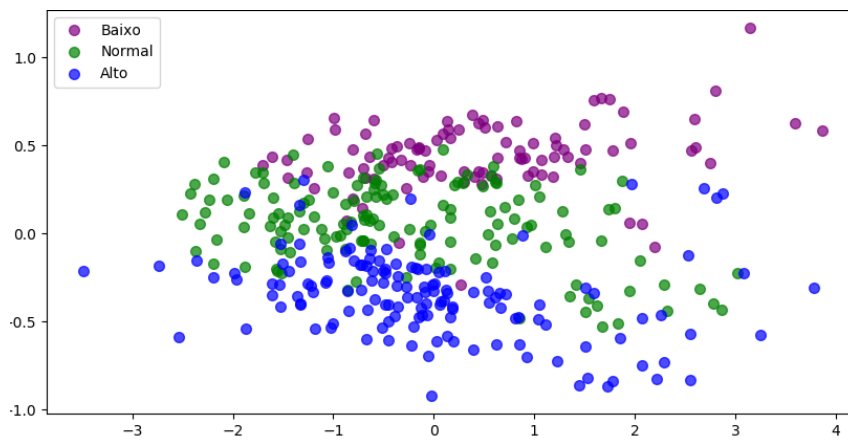


Figura G.83: Representação através de PCA dos níveis de referência do ácido úrico, baseada na média dos componentes RGB.

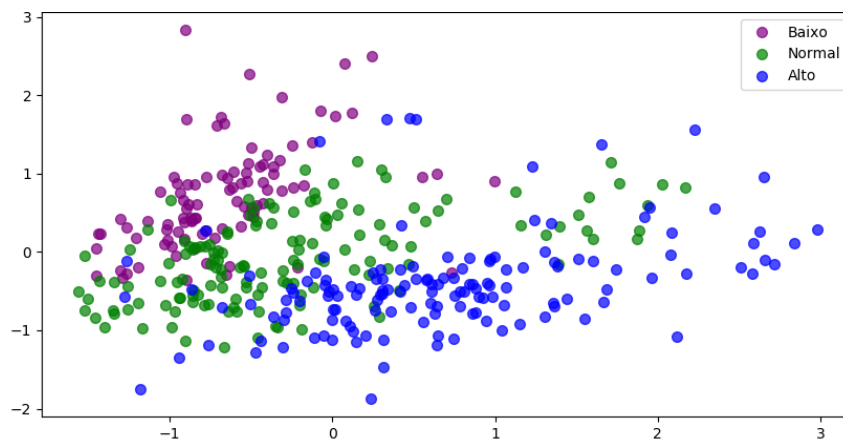


Figura G.84: Representação através de PCA dos níveis de referência do ácido úrico, baseada na média dos componentes HSV.

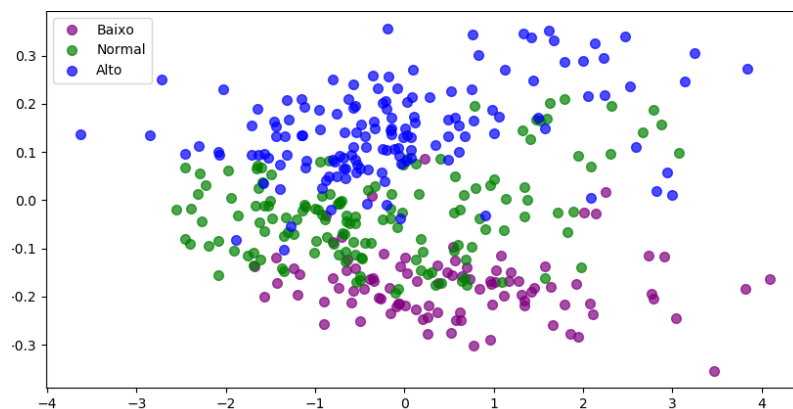


Figura G.85: Representação através de PCA dos níveis de referência do ácido úrico, baseada na média dos componentes XYZ.

APÊNDICE G. RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

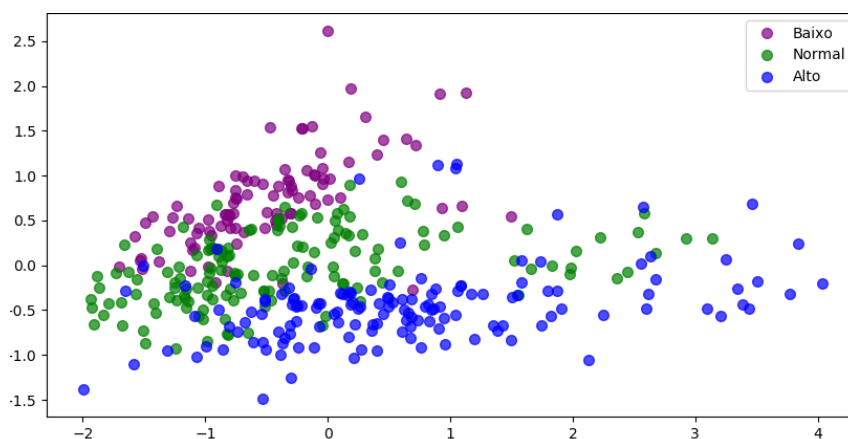


Figura G.86: Representação através de PCA dos níveis de referência do ácido úrico, baseada na média dos componentes HLS.

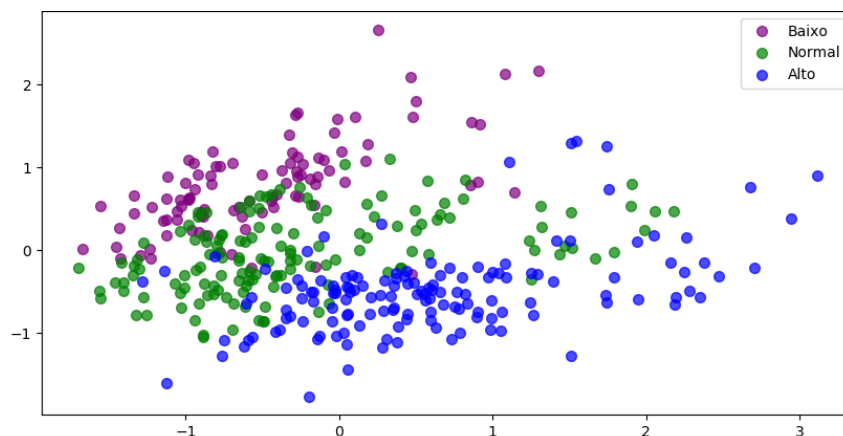


Figura G.87: Representação através de PCA dos níveis de referência do ácido úrico, baseada na média dos componentes LAB.

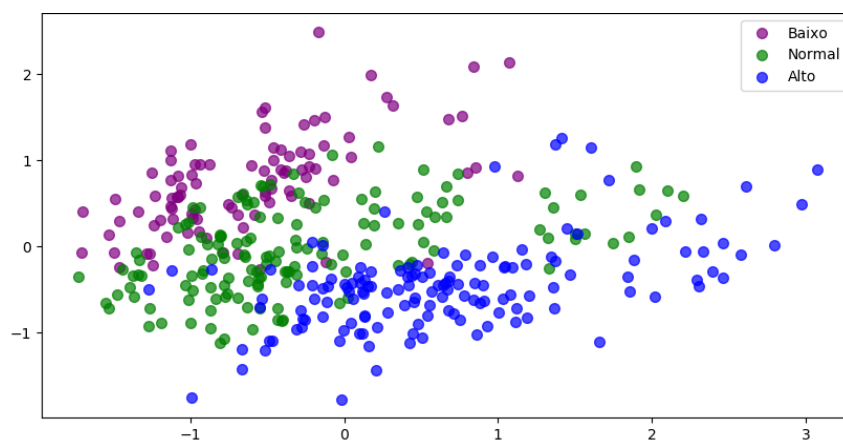


Figura G.88: Representação através de PCA dos níveis de referência do ácido úrico, baseada na média dos componentes YUV.

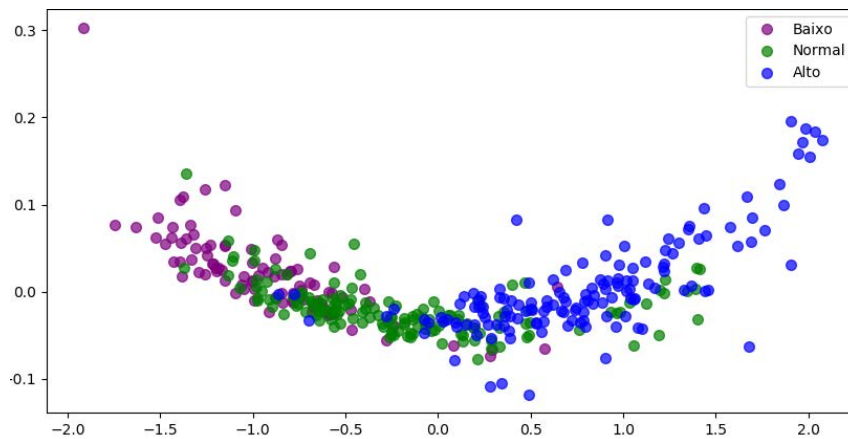


Figura G.89: Representação através de PCA dos níveis de referência do ácido úrico, baseada no rácio entre os componentes vermelho e azul do espaço RGB e componente H do espaço HSV.

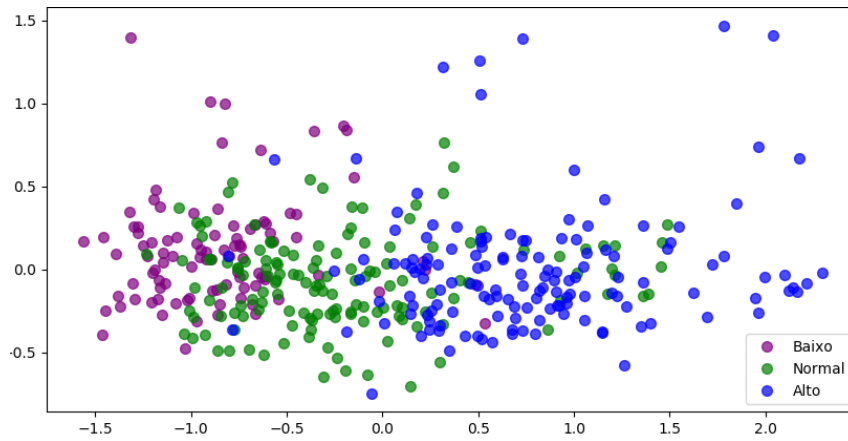


Figura G.90: Representação através de PCA dos níveis de referência do ácido úrico, baseada nos componentes a e b do espaço LAB.

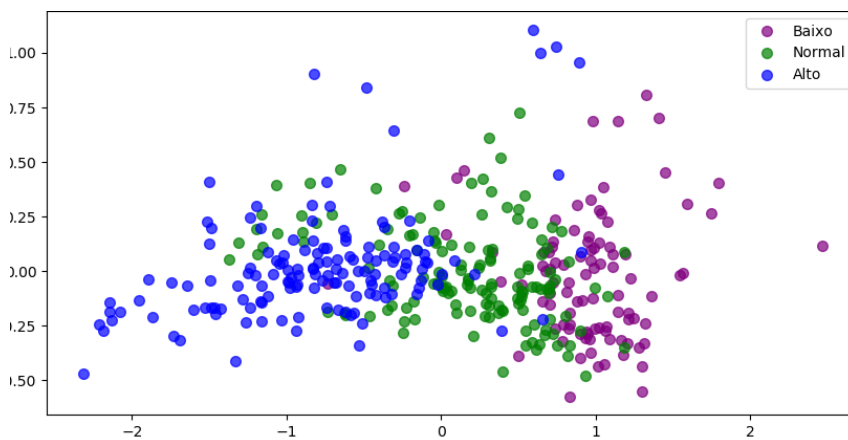


Figura G.91: Representação através de PCA dos níveis de referência do ácido úrico, baseada nos componentes H e S do espaço HSV.

APÊNDICE G. RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

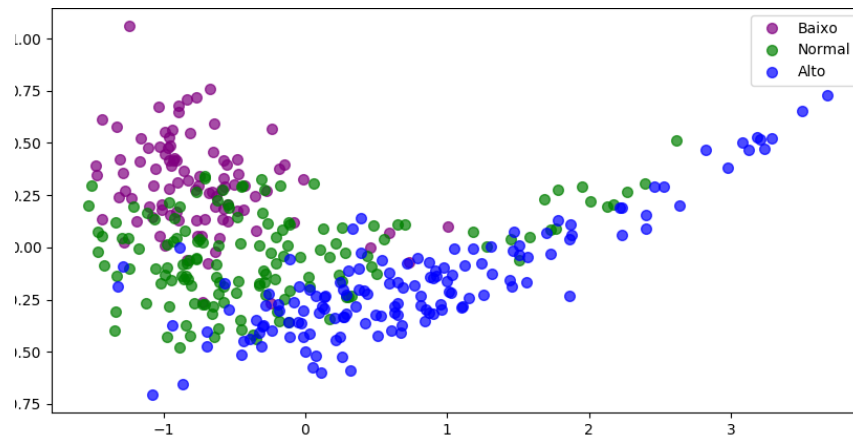


Figura G.92: Representação através de PCA dos níveis de referência do ácido úrico, baseada nos componentes H e S do espaço HLS.

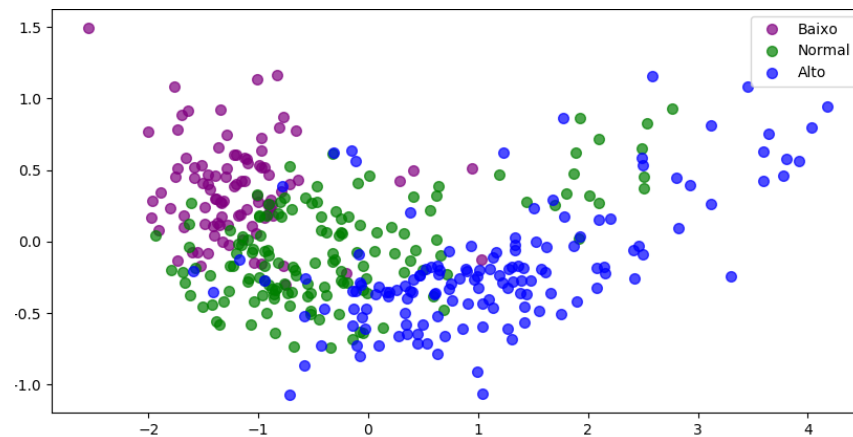


Figura G.93: Representação através de PCA dos níveis de referência do ácido úrico, baseada nos componentes H e S dos espaços HSV e HLS.

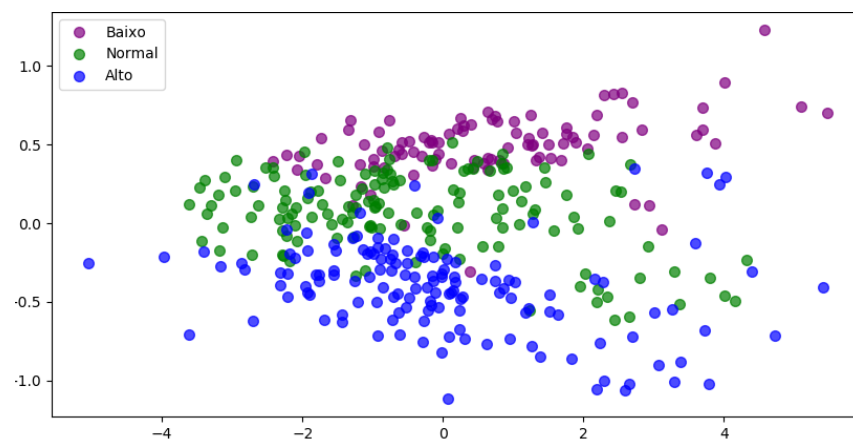


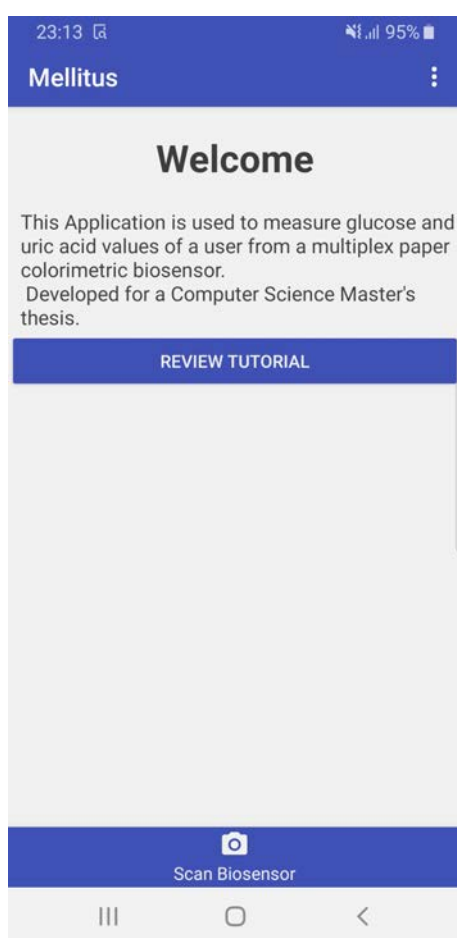
Figura G.94: Representação através de PCA dos níveis de referência do ácido úrico, baseada na combinação da média dos componentes do modelo RGB com as componentes V do modelo HSV, Z do modelo XYZ, A do modelo LAB e Y do modelo YUV.

## APÊNDICE

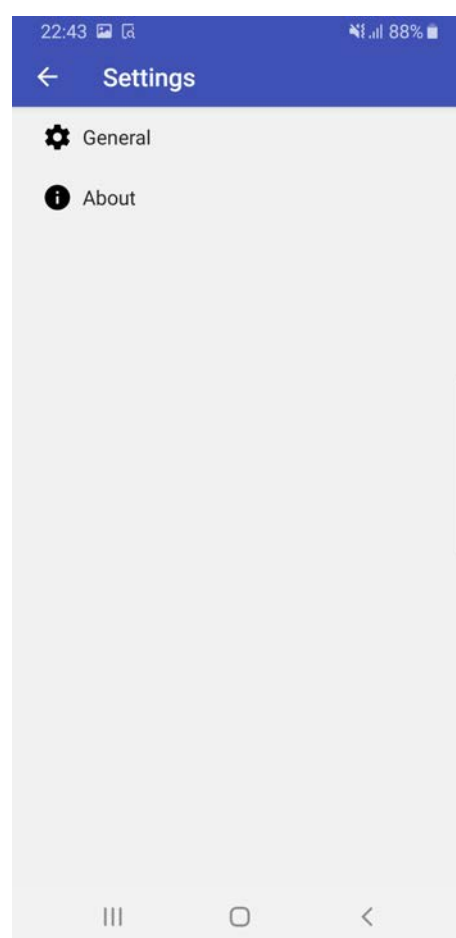
# H

## INTERFACE DA APLICAÇÃO

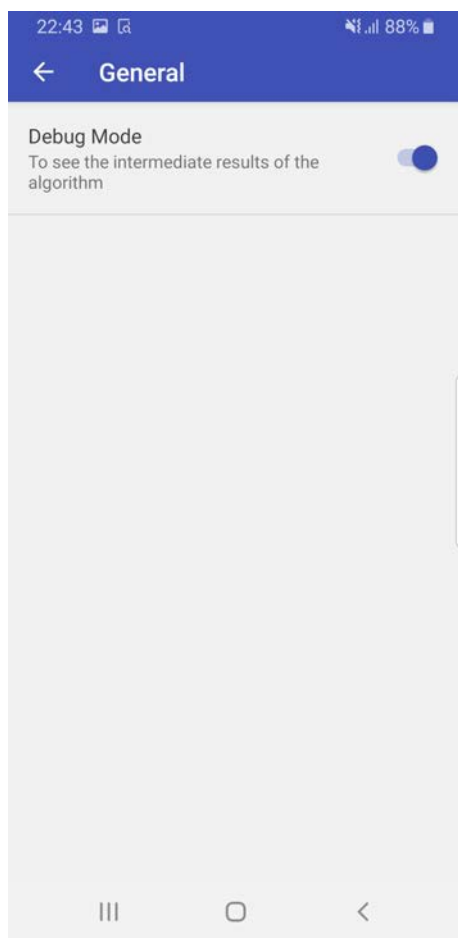
Neste apêndice estão disponíveis os ecrãs principais da aplicação móvel.



(a) Ecrã inicial



(b) Ecrã de preferências



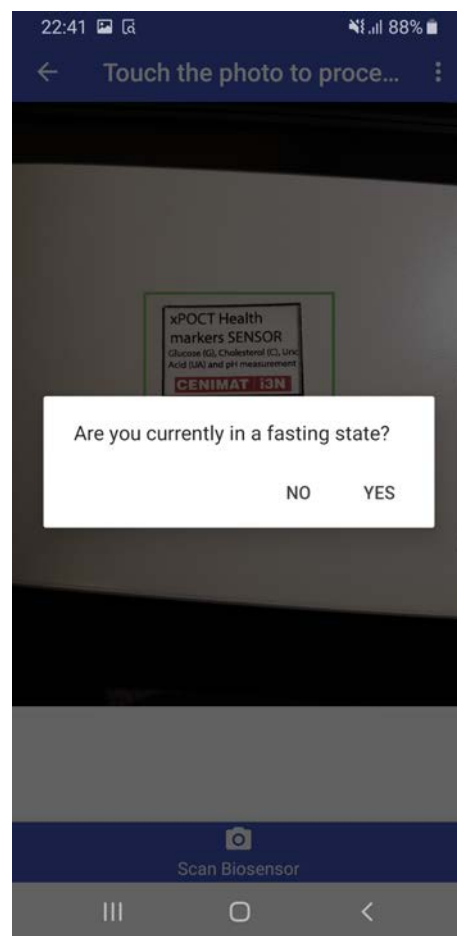
(c) Ecrã de preferências gerais



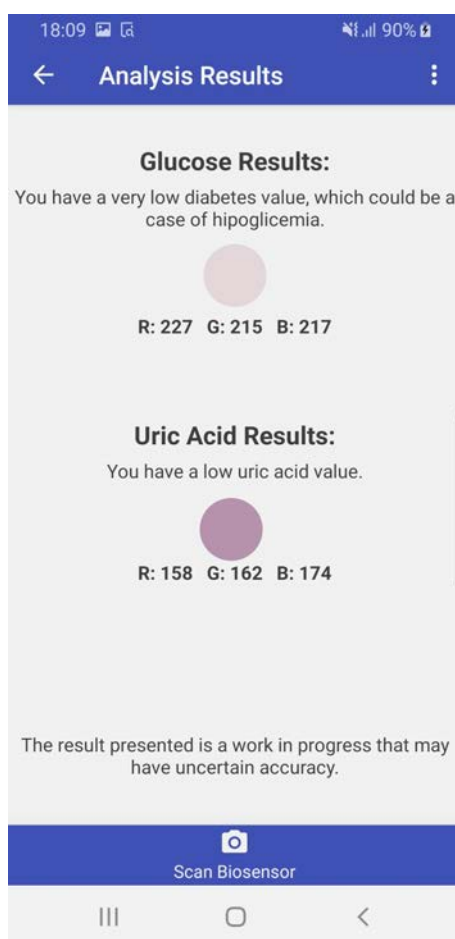
(d) Ecrã de deteção



(e) Ecrã do modo *debug*.



(f) Ecrã de escolha do estado da glucose



(g) Ecrã de resultados

Figura H.1: Ecrãs principais da aplicação móvel.



